NOTICE

This is the author's version of a work that was accepted for publication in Computers & Education. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in **Computers & Education 57 (2011), pp. 1971-1988, DOI: 10.1016/j.compedu.2011.04.007**. The final version was published on-line: **9-JUN-2011**.

Title page

1) Full title

Are educational computer micro-games engaging and effective for knowledge acquisition at high-schools? A quasi-experimental study

2) Authors

Cyril Brom, Michal Preuss

Faculty of Mathematics and Physics, Charles University in Prague

Malostranské nám. 25, 118 00, Prague, the Czech Republic

Daniel Klement

Institute of Physiology, Academy of Sciences of the Czech Republic

Vídeňská 1083, 142 00, Prague, Czech Republic

3) Source of funding

Development of the game was partially supported by the project CZ.2.17/3.1.00/ 31162 financed by the European Social Fund, the Budget of the Czech Republic and the Municipality of Prague. The research part of the project was partially funded by the research projects MSM0021620838, GA CR 309/09/0286, and MSMT CR LC554.

Are educational computer micro-games engaging and effective for knowledge acquisition at high-schools? A quasi-experimental study

Abstract

Curricular schooling can benefit from the usage of educational computer games, but it is difficult to integrate them in the formal schooling system. Here, we investigate one possible approach to this integration, which capitalizes on using a micro-game that can be played with a teacher's guidance as a supplement after a traditional expository lecture followed by a debriefing. The game's purpose is to reinforce and integrate part of the knowledge learnt during the lecture. We investigated feasibility of this approach in a quasi-experimental study in 70 minutes long seminars on the topic of animal learning at 5 classes at 4 different high-schools in the Czech Republic. Each class was divided to two groups randomly. After an expository lecture, the game-group played a game called Orbis Pictus Bestialis while the control group received an extra lecture that used media-rich materials. The time allotment was the same in both groups. We investigated the immediate and one month delayed effects of the game on students' knowledge reinforced and integrated by the game as well as on knowledge learnt during the expository lecture but *not* strengthened by the game. We also investigated students' overall appeal towards the seminar and its perceived educational value. Data from 100 students were analysed. The results showed that a) the game-playing is comparable to the traditional form of teaching concerning immediate knowledge gains and has a significant medium positive effect size regarding retention, b) the game-playing is not detrimental to information transmitted in the expository lecture but not strengthened by the game, c) perceived educational value and the overall appeal were high in the game group, nevertheless the perceived educational value was slightly lower in the game group comparing to the traditional group. Our results suggest that the proposed approach of harnessing educational computer games at high-schools is promising.

Keywords

interactive learning environments; multimedia/hypermedia systems; simulations; secondary education; applications in subject areas

1. Introduction

Nowadays, educational computer games, also called serious games, are used in multiple fields, including military training, medical and public health training, rehabilitation, and foreign language practising (e.g., de Freitas, 2006). Many have argued that they could also support classical curricular schooling. In terms of Bloom’s taxonomy of learning outcomes (Bloom, 1956), the main arguments of the proponents of computer games based learning concentrates on cognitive and affective elements. With respect to the affective domain, some argue that playing computer games, as a part of curricular teaching, can substantially increase motivation of learners (e.g. Barab et al., 2005; Kirriemuir & McFarlane, 2004). This claim is often supported by the fact that computer games, and games in general, create *intrinsic* motivation through fantasy, control, challenge, curiosity, and competition and/or cooperation (Cordova & Lepper, 1996; Malone, 1981). Concerning the cognitive domain, the main proposal is that employing computer games, and interactive digital simulations, as educational tools may help for developing advanced knowledge and skills, and generating deeper understanding of certain key principles of given topics, mainly when dealing with complicated and multifaceted issues that are hard to comprehend through factual knowledge only. Complex computer games immerse players in rich and stimulating environments, allow them to explore numerous strategies for action and decision, and require them to complete demanding tasks with increasingly difficult objectives (Facer et al., 2007). Thus, they could help the learners to acquire mental models of complicated processes. Immersive learning environments created by games and simulations also have the potential to engage students in scientific practise (e.g., Rivers & Vockell, 1987; Aitkin, 2004; Squire & Jan, 2007; Nilsson & Svingby, 2009). Additionally, complex games could promote general problem-solving skills, goal-oriented behaviour, and, in cases of multi-player games, social networking (Gee, 2003; de Freitas, 2006; Squire, 2005; Sandford et al., 2007). Computer games could also help to develop strategic thinking, group decision-making, and other higher cognitive skills (Arnseth, 2006; de Freitas, 2006). Unfortunately, these claims have not been supported by sufficient empirical findings in the context of curricular education. Only a limited number of empirical studies exist and some of them presented mixed results (as will be reviewed below in more detail).

One important issue reported repeatedly is that integration of an educational computer game with formal schooling environment is difficult, arguably more difficult than in many other contexts, e.g. in workplace. In general, educational computer games pose new challenges for the educational system (e.g., Egenfeldt-Nielsen, 2005; Sørensen et al., 2007; Becker, 2008; Sisler & Brom, 2008). Many practical barriers for their integration exist, ranging from the unintelligibility of interfaces and game rules for some teachers and non-players, to a lack of access to equipment, e.g. up-to-date video cards, to barriers posed by fixed lesson times (e.g., Squire, 2004; Egenfeldt-Nielsen, 2005; Sisler & Brom, 2008; Klopfer, 2008; Ketelhut & Schifter, 2011). Therefore, researchers and designers have been investigating various approaches to include educational computer games in regular curricula at schools.

In the present study, we investigate one approach of incorporating a computer game into curricular education and we assess its effectiveness. Extending the older idea of computer-based simulations (e.g., Lee, 1999), this approach capitalizes on using a computer micro-game that can be played with a teacher's guidance as a supplement after a traditional expository lecture and then followed by a short debriefing. We define operationally micro-games as relatively simple computer games that do not require special skills to play and that challenge players with clearly defined goals reachable within minutes or tens of minutes of game-play. Moreover, the learning is an integral part of the playing and vice versa as opposed to the classical "drill-and-practice" edutainment software. The educational objective of such micro-game, in our type of usage, is to *reinforce* and *integrate* (in terms of Thomas & Hooper, 1991) part of the knowledge learnt in the expository lecture. Thus, this approach focuses on acquisition of the core knowledge of a particular topic as opposed to facilitating high-level skills like communication abilities or scientific problem solving skills.

We chose high-school students as the target audience since this is arguably the most ignored segment concerning educational games (9^{th} till 12^{th} grade in the US system). Only few educational computer games, or instructional computer-based simulations, have been developed for high-school students. For instance, in the review of Lee (1999) surveying 19 comparative studies of effectiveness of "computer-based instructional simulations," only 5 were targeted at high-school audience; and of the 48 empirical research articles on effectiveness of computer as well as non-computer "instructional games" reviewed by Hays (2005), only 4 were intended fully or partly for high-school students or students of that age. This is in line with observations of Egenfeldt-Nielsen (2005; p. 93) that general edutainment is strongest in pre-school and early-school area, and of Rieber (1996) that games have the greatest acceptance in the early grades but teachers' and parents' interest in their use declines in middle and secondary schools (but see also de Freitas, 2006). Moreover, adolescents are more difficult to engage in school learning and harder to motivate than younger children (e.g. Eccles & Midgley, 1989; cited from Papastergiou, 2009). Recently, a few studies regarding learning outcome of educational games for high-school students have been published; the most relevant of them for our study will be reviewed below.

To assess the effectiveness of our approach, we have stated several hypotheses related to knowledge gain as well as to several affective variables. We now proceed with a review of general findings from older literature on gaming in curricular education based on which we formulated our approach and the hypotheses. Then, we continue with review of recent studies on integration of computer games in the formal schooling systems, focussing predominantly at high-schools and quantitative outcomes. Similarly to (Randel et al., 1992), we exclude from our present interest business games since they do not cover traditional high-school subjects (see Faria, 2001 for a review). We also do not discuss “drill-and-practise” and “quiz-based” software here (see, e.g., Niemiec & Walberg, 1985; Papastergiou, 2009) as it is not part of our definition of micro-game. After the reviews, we introduce hypotheses for the present study.

2. Background to the study

2.1 General findings from older reviews of educational game studies

The effectiveness of game-based activities in the curricular education (and elsewhere) has been already studied in past. The most relevant activities for present purposes are non-computer experiential simulations, interactive digital simulations and educational computer games. Are these activities effective?

General reviews of studies comparing instructional effectiveness of games to more conventional form of instruction reported mixed results (e.g., Randel et al., 1992; Hays, 2005). The main conclusions of these two reviews are that a) in most studies investigating motivation, learners reported more interest in simulation/game activities than in the conventional instruction but that this is not necessarily linked with better learning; b) in studies investigating cognitive performance[1] immediately or shortly after the treatment, games were usually at least as effective as other kinds of instruction but only rarely better (Randel et al. presents more optimistic picture than Hays); and c) that games can be detrimental to learning if they do not include instructional support (Hays, 2005; p. 47). These reviews also included information from earlier reviews of comparative studies of simulation gaming, that is, mostly of experiential, non-computer games (Peirfy, 1977; Bredmeier & Greenblat, 1981; Dorn, 1989). Conclusions of the older reviews were similar.

Concerning long lasting effects of the game-based learning on cognitive performance, information about retention is more useful than information about the immediate effect. In this regard, Pierfy (1977) made one interesting point: simulation games might improve retention, as measured by delayed post-tests, despite that no immediate effect is found. Eleven studies surveyed by Pierfy assessed retention, of these, eight showed retention significantly better with a simulation/game and three showed no difference (cited from Bredmeier & Greenblatt, 1981). More recently, retention has not been investigated much, though Randel et al. (1992) added little additional evidence to support the Pierfy's point.

---

[1] Note that studies investigated whether playing improve various cognitive abilities, ranging from high-level skills to mastering core knowledge of a particular topic. The term "cognitive performance" is used here as an umbrella term covering all of these.

In a review on “computer-based instructional simulations,” Lee (1999) reports that students using a simulation generally outperformed students from the control group in terms of “academic achievements,” particularly when the simulation was supplemented by expository information. It is impossible to determine with certainty from the Lee’s review whether the achievements were measured by immediate or delayed post-tests, but the former is more probable. Five studies surveyed by Lee also investigated students’ attitude. Students showed negative or very little preference to simulations, which does not accord well with the findings mentioned above.

Additionally, debriefing is known to be an important phase in using simulation games (e.g., Peter & Vissers, 2004) and some suggest that it should be also used in computer games instructional programs (e.g., Hays, 2005). Finally, there is evidence that minimal or no guidance is problematic in general (e.g. Rivers & Vockel, 1987; Kirschner et al., 2006; Tobias & Fletcher, 2007).

Thus, based on the older literature, we can conclude that when constructing an instructional approach to using educational computer games at high-schools, a promising model can be a) to include also other instructional support, i.e. to use the game as a supplement rather than the main or even stand-alone activity, and b) to include debriefing. We can expect that learners will demonstrate higher knowledge gain in retention tests, but not necessarily immediately after the lesson. However, the data on retention are limited. We may also speculate that students may demonstrate higher overall appeal towards the course including a computer game-playing activity, but this is also not very clear.

Does this picture accord well with more recent findings on educational computer games?

2.2 Integration and effectiveness of modern computer games in curricular education

As already said, researchers and designers have been investigating various approaches to integration of modern educational computer games in schools. A common starting point is development of a game tailored directly for the purposes of curricular schooling as opposed to entertainment, as it turned out that usage of commercial entertainment off-the-shelf games (COTS) for teaching and learning is particularly problematic (Squire, 2004; Egenfeldt-Nielsen, 2005; but see also de Freitas, 2006).

One possible integration strategy is to design a several hours or days long seminar with various activities organised around playing the game, by-passing the problem with fixed lesson duration. At least two high-school oriented empirical studies (Buch & Egenfeldt-Nielsen, 2006; Brom et al., 2010) suggested that this approach is promising, reporting that students demonstrated positive attitudes towards the game and that the majority of players claimed that they learned more or at least as much as they usually did. However, both studies were uncontrolled and did not investigate real learning outcomes.

Other method is to employ PDA devices dedicated only to game-based classes, by-passing problems with a computer lab. A prominent proponent of this approach is Klopfer (see Klopfer, 2008, for a summary). While most results of his team are promising both in affective and cognitive terms and produced valuable insights into actions, strategies and learning forms taken and/or developed by the players, and also into what students can actually learn from the games, he generally employs uncontrolled design and tend to use descriptive reports in his studies (see also Barab et al., 2007; Squire & Jan, 2007; Nilsson & Svingby, 2009 for other similar studies). It is important to complement qualitative findings with quantitative outcomes of controlled studies.

Yet other possibility is designing a game as a home-play activity (see again Brom et al., 2010) or out-of-school activity (e.g., Rosenbaum et al., 2007; Wrzesien et al., 2010; Huizenga et al., 2009). The latter three studies employed both qualitative and qualitative measures, but only the second and the third were controlled. Concerning the affective domain, the first and the second study reported positive results while the third one mixed results, most likely due to technical problems (this was a mobile game requiring a complex technology). Concerning the cognitive domain, all studies reported some learning gains, but in the second one, the game-based activity did not present statistically significant differences with the traditional type of class, and in the third, it did, but the control group received a shorter treatment than the experimental group. The first two studies did not investigate retention, the third employed one-week post-test but no immediate post-test (Huizenga, personal communication[2]).

Finally, one can employ micro-games. The main idea behind this approach is that they can be played easily on older school computers within school lessons with fixed duration and that their game-play can be designed easily to match parts of the curriculum. Descriptive studies of Wilensky's team of various simulations developed by Netlogo modelling toolkit (Wilensky, 1999) suggest that this approach is promising (e.g., Wilensky & Novak, 2010). Annetta et al. (2009) investigated a high-school micro-game on genetics and found significant increase in students' level of engagement while interfacing with the game but no learning effect comparing to the control group receiving usual genetics instruction. Ioannidou et al.

[2] An e-mail from 26th September 2010.

(2010) investigated a collaborative micro-game, called "collective simulation" in that work, on the topic of cardiopulmonary system. Learning gains were statistically significant in favour of the game group, but only on a subset of questions that were arguably mostly related to the game activities, as opposed to the rest of the questions, on which only negligible positive effect was found. The rest of the questions was "collected from various conventional assessments" (p. 158) and therefore may have not always tested what the game taught. These studies focused on short-term effects of the games (e.g., Annetta et al. (2009) administered questionnaires 4 days after the treatment (personal communication[3])). Finally, Wong et al. (2007) investigated a game on physiology concepts for undergraduate college students. There was knowledge gain when comparing the game group with a text-only group in the immediate post-test as well as in the follow-up test one week later. Results also favoured the game group comparing to the text-only group with respect to several motivational variables. Wong et al.'s study was the only one we were able to find that explicitly investigated also retention, except of the several studies from seventies mentioned by the reviews above and the study of a complex COTS game used in a several weeks long, high-school history seminar (Egenfieldt-Nielsen, 2005), which found no gain comparing to the control group in the immediate post-test but a small gain in the five month follow-up test.

Thus, the findings are, more or less, in line with the older findings on more general educational game-based activities; a) educational computer games seem to be motivating enough when technical problem are overcame, b) students demonstrate some knowledge gains, but in controlled studies, the games are only sometimes more effective than other forms

---

[3] An e-mail from 30th October 2010.

of instruction if measured immediately or shortly after the treatment. Additionally, c) the data on retention are thin. The study of Ioannidou et al. (2010) further suggests that d) the effectiveness in the cognitive domain may depend on whether we measure knowledge directly taught by the game as opposed to more general knowledge, e.g. to which students are exposed in the supplementary lecture (in fact, this point is not very surprising).

3. Research questions

Based on the abovementioned findings, we set the following hypotheses for our present study:

1) The game will be at least as effective as traditional instruction concerning the gain of knowledge that the game is supposed to reinforce and integrate, if measured immediately after the exposure.

2) The reinforced and integrated knowledge will be retained better by the game group.

3) The game will not be detrimental to knowledge that was transmitted by the expository lecture but not reinforced and integrated by the game; neither immediately nor later.

4) The course with the game will be more engaging than the lecture-only course.

We focus on quantitative outcomes as we are mainly interested in the overall effectiveness of the game. It would be also interesting to investigate *how* students learn from the game and what is the difference between students' learning styles when exposed to the game versus traditional instruction (using more qualitative approach), but this is out of our present scope.

Even though we make no explicit hypothesis regarding gender differences, we also analyse our data with respect to gender. Such analysis has been only rarely performed in the past even though the issue of possibly different impact of educational games on boys and girls is important. Thus, our analysis can serve as a base for future studies.

To test these hypotheses, we have developed a single-player micro-game called Orbis Pictus Bestialis on the topic of animal learning. This topic is part of the Czech national curricula, but it is notoriously known as difficult. In fact, the game is part of a larger project on development (and/or translation to Czech) of educational computer games for high-school students. One long-term goal of this project is to pinpoint design principles contributing most to success of educational computer games (in the context of secondary education). Thus, this study can be regarded as a first step towards achieving this goal, a pilot study.

As some past reviews of educational game-based activities complained that the activity being studied is not described in sufficient detail (see, e.g., VanSickle, 1986, p. 247; Hays, 2005, p. 48), complicating drawing more general conclusions or making it impossible to conduct a rigor meta-analysis, we now proceed with a detailed description of our game.

4. Orbis Pictus Bestialis: Game description

The micro-game Orbis Pictus Bestialis (OPB) has been designed to reinforce and integrate (in terms of Thomas & Hooper, 1991) part of the knowledge learnt in the expository formal instruction. Thus, the traditional lecture *plus* the game should be viewed as a unit; the game is not a stand-alone activity.

The educational objective of this unit is to teach students basics of ethology, behaviourism, and animal learning. Part of this topic is so-called *positive reinforcement learning*, which is a method frequently used to train real animals. The game *per se* allows students to explore major phenomena of positive reinforcement learning, to practice basic training techniques such as behaviour capturing, shaping, and chaining and to learn how to use a clicker during the training. Thus, the game has been designed to reinforce and integrate *this* knowledge, but not other more general knowledge taught during the lecture (see Appendix A for details).

OPB is a single-player game and it has three levels. In each level, the student has to train an animal to perform a task; namely to train a dog waving one front leg on a verbal stimulus, to train a lemur entering into a transportation box and closing the door behind it, and to train a parrot speaking. Only the first level is employed in this study.

Each virtual animal is underpinned by a biologically plausible behavioural model developed with experts on animal learning, and the course of learning of the animals reproduces qualitatively the course of learning of their real counterparts. The model is based on Q-learning (Watkins, 1989), which is a well-known machine learning algorithm. The model exhibits various basic learning phenomena, such as forgetting and sensitiveness to timing of reward (but some advanced phenomena are not modelled, such as consequences of various kinds of reward scheduling). The animal's state is represented by real-valued variables, offering a player a large state-space for exploration. The learning happens in real time and it is often quicker than it would be in reality (due to relatively short time allocated for the play).

In the first level, a player interacts with a dog which spontaneously performs several different actions; wagging the tail, barking etc (Fig. 1). The player can present seven stimuli to the dog by pressing buttons: giving a food, producing a sound with a clicker, touching the dog's front paw by the trainer's hand, raising the trainer's hand to three different positions in front of the dog, saying the word "PAC". In general, any action performed by the dog can be reinforced. As a consequence, the probability of this action will increase - this is operant conditioning. Additionally, a stimulus can be associated with another stimulus - this is classical conditioning - or an action can be brought under a stimulus' control, i.e. performed by the dog only after the presence of the stimulus. An important part of the game is the usage of the clicker. The clicker makes it possible to precisely mark the end of the desired action. In this respect the clicker is superior to the food. However, before the clicker can be used, its sound should be associated with the food.

Two training methods can be practiced within the first game's level: capturing and shaping. Capturing: the trainer waits until the dog produces the desired action; after the action has been completed, the trainer reinforces the action either by giving the food or by using the clicker. Shaping: sometimes the desired behaviour is so uncommon that the dog never exhibits it spontaneously. In this case the player can reward behaviours which are similar to the desired behaviour and which spontaneously occur with a reasonable frequency. As the training proceeds, the player gradually increases requirements for the animal.

The goal of the student is to train the dog waving one front leg on a verbal stimulus. In this task, the student practices the following principles: forming an association between stimuli, capturing behaviour, shaping behaviour, bringing behaviour under a stimulus control, timing of rewards and using the clicker during training. The optimal course of the training is as follows:

1) Teaching the dog that a certain sound (the clicker) precedes the reward (food; this is the formation of association between two stimuli - Fig. 2).
2) Rewarding the dog when the trainer's hand touches its paw.
3) Holding the hand slightly above the dog's paw and rewarding the contact of the hand and the paw (players capture the dog's behaviour as they wait until the dog performs the action).

4) Raising the hand increasingly higher, forcing the dog to raise its paw higher (the shaping process – Fig. 3).
5) Holding the hand so high up that the dog cannot reach it when sitting, and rewarding the exact instant (by the clicker) when the paw is in the desired height just before the dog starts jumping (the clicker terminates the behaviour, the dog learns that lifting its paw is rewarded, not the contact between its paw and the trainer's hand - Fig. 4).
6) Pronouncing the verbal command just before raising the hand (bringing behaviour under stimulus control – Fig.5).
7) Pronouncing the verbal command without raising the hand (this is a continuation of the preceding point; the dog should lift up its paw after the player says the verbal command only).

An expert can teach the virtual dog this task in about 5 minutes. The challenge is that exact sequence of steps is not given (due to the plausible computational model of learning). It is necessary to repeat most of the steps several times, sometimes regressing a step or two back. The dog exhibits the tendency to perform not the desired action but a similar one that is simpler. The first step of the training procedure may be performed later on. Many sidewalks are possible; for instance, a student can teach the dog wagging the tail on a stimulus. Timing of the reward is crucial. With repeatedly wrong reinforcing actions, the student can confuse the dog so that it will be almost impossible to finish the task any more (as is the case in reality).

From the theoretical perspective, the important point is that the internal state of the dog is not visible to the player. The only feedback the player gets is the actions the dog performs. While this limited feedback may be disputable from the gaming perspective (cf. Malone, 1981), it plausibly models what kind of feedback the student would get in reality. Teachers' guidance is important for this reason.

--- Insert Figure 1 about here ---

--- Insert Figure 2 about here ---

--- Insert Figure 3 about here ---

--- Insert Figure 4 about here ---

--- Insert Figure 5 about here ---

4.1 Related work

Educational computer games, simulations and simulation-based activities on a biological topic are few comparing e.g. to games for teaching math or physics (see, e.g., the reviews of Randel et al., 1992; Hays, 2005 for exact numbers). Most biological games/simulations address the topic of population modelling and/or predator-prey cycles (e.g., Lutterschmidt & Schaefer, 1997), genetics and/or evolutionary theory (e.g., Rivers & Vockel, 1987; Church & Hand, 1992; Klopfer, 2008; Wilensky & Novak, 2010; Annetta et al., 2009), plant biology (e.g., Rivers & Vockel, 1987; Mikropoulos & al., 2003), physiology (e.g., Wong et al., 2007; Ioannidou et al., 2010), disease spreading (e.g. Klopfer, 2008), or virtual dissections (e.g., Akpan, 2001).

We were able to identify only two computer games/simulations on the topic of animal learning: Pavlov's Dog (NBP, 2001) and Sniffy (Alloway et al., 2005). The former one is a simple Flash game on the topic of classical conditioning only, tailored for primary school audience. In this game, learning seems to be only imitated, that is, there is most likely no underlying model and all the reactions are pre-scripted giving us a simple state space that can be described by a finite-state machine. On the other hand, Sniffy is a complex simulation of a rat behaviour that models more learning phenomena than OPB and is mainly suitable for advanced high-school and university courses. OPB lies in between these two applications in terms of complexity; it is most suitable for regular biology classes at high-schools or middle-schools. Additionally, in OPB, the animals are intentionally situated in environments familiar to students, e.g. in a zoo or in home, and they are being taught tasks the students may already know about. Contrary, in Sniffy, the rat is taught experimental tasks in a laboratory environment. Because of the unfamiliarity of that context to most high-school students and because of the complexity of the Sniffy's user interface, we speculate that using Sniffy in regular high-school class may increase students' cognitive load too much, especially should Sniffy be used only over a short period, e.g. 20 minutes. Besides Sniffy and Pavlov's Dog, several commercial games on the topic of animal care exists (e.g., Nintendogs, World of Zoo), but they feature none or very limited learning, which makes them unsuitable for purposes of teaching the topic of animal learning.

Most of the abovementioned games/simulations were not evaluated using controlled design with retention tests. Of interest is here a study (Venneman & Knowles, 2005) of benefits of using Sniffy in an undergraduate course as a supplemental teaching tool to present schedules of reinforcement in operant conditioning (which is an advanced subtopic of animal learning, see, e.g., Blackham, 1974). The experimental group outperformed the control group in the cognitive domain. Because students used the simulation as homework during the term, post-tests most likely measured *retained* knowledge. This adds additional merit to our second hypothesis.

5. Method

5.1 Experimental design

The study compares a teaching session consisting from an expository lecture followed by an educational micro-game on animal learning (game group) against a teaching session consisting from the same expository lecture followed by another lecture that used media-rich materials (traditional group). The game as well as the additional lecture was aimed at reinforcing and integrating part of knowledge learnt in the introductory lecture. Thus, the objective of the additional lecture in the traditional group was the same as the objective of the game (see Section 5.3 for details).

The study was set in five classes in four high-schools in cities of Prague and Kladno in the Czech Republic. At high-schools in the Czech Republic, it is typical that theoretical lectures are supplemented by practical seminars; often, one school hour (45 minutes) is devoted to the theoretical part and another (45 minutes) to the practise. The two school hours are often consecutive. The study took place within such 90 minutes long block. To avoid practise effect and cueing of students on what should be remembered, we employed post-tests only design (see e.g. Egenfeldt-Nielsen, 2005 for a similar design; see also Randel et al., 1992; Judd et al., 1991). Importantly, we administered not only an immediate post-test but also one month delayed post-test. The time allotment was, both for the game and the traditional groups, as follows:

- 5 minutes: introduction, explaining the purpose of the study (without mentioning the hypotheses);
- 40 minutes: a theoretical lecture;
- 20 minutes: either interaction with the game with the teacher's guidance (game group), or an extra lecture (traditional group);
- 5 minutes: debriefing; in both groups;
- 20 minutes: a post-questionnaire; in both groups.

Fig. 6 outlines the experimental design.

To control for the lesson's content, all lessons were given by the same person, namely the first author of this paper (C.B.), who is well familiar with the topic of animal learning. Another experimenter (M.P.) responsible for the technical part (installation of the game etc.) and one regular teacher of the respective class were present.[4]

The retention post-test was administered in both the control and the treatment groups as part of a regular lesson (this took about 15 minutes). The students were not informed in advance about the retention post-test.

---

[4] Some authors (e.g., Hays, 2005, p. 48) find it problematic when main authors of a game evaluate their own product. We tried to address these concerns. The main authors of OPB are the second author of the paper (M.P.) and another person that was not involved in the research part. The lecture was given by C.B. and the data analysis conducted by D.K., whose contributions to the game design was only modest.

--- Insert Figure 6 about here ---

5.2 Participants

Together, we recruited four high-school classes in Prague and one in Kladno, the Czech Republic. Every class was divided to the game and the traditional group on a random basis. However, the teacher in the Kladno's class preferred that the part that should have served as the traditional group should receive the treatment (i.e. the game) as well. Though this was an encouraging moment, we did not include this class in the present paper. Additionally, we excluded from the evaluation 17 students that had not finished either the immediate or the retention post-test.

Thus, we analysed data from 100 students from 4 classes from 3 high-schools, all of which were located in Prague, the Czech Republic. The numbers of students according to gender, school/class and intervention are presented in Table 1. All the classes are from high-schools that are rated as "above average" by the Czech School Inspectorate[5]. The two groups were similar in the distribution of gender (Table 1). To anticipate, the results also shown that the groups were similar in the age as well as in other biographical variables related to their computer literate (Table 2). Thus, we have a relatively homogenous sample of students.

--- Insert Table 1 about here ---

---

5 The schools were rated in 2009 (School A), 2010 (School B), and 2008 (School C); www.csicr.cz, accessed 17.7.2010.

5.3 Materials

The general aim was to compare a micro-game against an extra lecture that has a traditional format. Though the situation is changing in the Czech Republic, most lectures at high-schools are still given by means of the traditional frontal method using chalk and blackboard. To avoid the effect of mere multi-media exposure, we decided to supplement all our formal lectures with power-point presentations with text and colour figures. Additionally, the expository lecture featured 3 approximately 2 minutes long videos on the topics of a) Konrad Lorenz and imprinting, b) Pavlovian conditioning, and c) operant conditioning of a pigeon in a Skinner box. The extra lecture in the control group employed an extra video on the topic of machine learning using reinforcement techniques (see also below). Thus the control group was exposed to four videos in total while the game group to three videos. The content of the lectures was prepared with experts on animal training.

After the expository lecture, the students assigned to the game group were instructed to move at computers and run the game. The interface and the control of the game were explained briefly and the goal of the first game's level introduced. The game was presented as a competition; the first two students achieving the goal, that is, training their virtual dog waving one front leg on a verbal stimulus, won a poster. This actually added an aspect of extrinsic motivation, though our informal observation is that the students were much more interested in the game as such than in the posters. The teacher (C.B.) continuously commented what happened in the game (typically, only few students listened to each comment due to playing the game, thus many comments were repeated).

The students assigned to the traditional group listened to the extra 20 minutes long lecture (see Fig. 6). Like the game, the extra lecture was designed to reinforce and integrate the knowledge on major phenomena of positive reinforcement but not other topics introduced during the expository lecture (see Appendix A). Namely, the course of training a dog waving its front leg was explained in detail. There are more ways how to train the dog to perform this task; only the way modelled by the game was introduced using the in-game graphics, i.e. figures such as those on Figs. 2 - 6. Then the basics of how positive reinforcing works on experimental animal *Aplysia californica* at the neural level was explained (see, e.g., Bear et al., 2007). Finally, it was explained that robots and virtual agents can be also trained using positive reinforcement methods. That was supplemented by the fourth extra video.

5.4 Questionnaires

Two questionnaires were constructed by the first author of the paper (C.B.) and a psychologist: a) an immediate post-test, b) a one-month delayed post-test. The immediate post-questionnaire comprised three parts: i) eliciting biographical and ICT-literacy data, ii) addressing students' engagement in the lecture in general and the game in particular, including feedback on the game's interface, iii) an open response knowledge test. The retention test contained only Part (iii).

Only the most relevant data are presented in this paper. Concerning ICT-literacy, the most important questions are those in which we asked students to specify on a 4-point ordinal scale how often they use a computer (1: "less than 1 hour a week" - 2: "1-5 hours a week" - 3: "6-10 hours a week" - 4: "more than 10 hours a week") and how often they play computer games (1: "never" - 2: "less then 1 times a week" - 3: "once a week or more but less than every day" - 4: "daily"). For the engagement/feeling questions, we used 1-4 Likert-like scale. The most relevant questions are "How did you like the whole seminar?", "Do you think that you have learnt something?”, and “How did you like the game in particular?” The last question was included only in the game group’s questionnaire.

To our knowledge, there is no standardised knowledge test on the topic of animal training in the Czech Republic, therefore, we created the immediate and the retention open response knowledge tests. Both had 8 questions and were created by the first author of the paper together with an animal training expert and a psychologist with respect to two criteria: they should correspond to tests which are usually given to students at high-schools and they should cover topics presented in the seminar, i.e. contain both general questions on the topic of animal learning (Appendix A, Points 1-3) as well as questions focussing on the positive reinforcing (Appendix A, Point 4). An additional requirement was that the questions in the retention test must be similar to the questions from the immediate test but not exactly the same. Our intention was not to compare exact scores from the retention test against the immediate test but to use the scores for group comparison only. Nevertheless, we wanted to achieve similar level of difficulty in both tests and to cover the same topics.

Two independent experts were asked to divide these questions into two groups assessing either the topic of animal learning in general (“animal learning question”; G1-questions) or positive reinforcing in particular (“positive reinforcing” ; G2-questions). A complete agreement was required. The two groups of questions were planned in advance. For the immediate post-test, 3 questions were assigned to G1-group and 5 to G2-group. For the retention test, the assignment was 4:4. No question was rated as ambiguous.

Two persons, the last author of this paper (D.K.) and another animal expert, independently of each other scored the student’s answers on a 3-levels ordinal scale as “incorrect answer” (0 points), “incomplete answer” or “partially correct answer” (1 point) and “correct answer” (2 points). The points given by each scoring person were averaged across each group of questions and then averaged across the two scoring persons. As a result the performance of a student in one test was represented by a number ranging from 0 (all the questions within the group of questions were answered incorrectly according to both scoring persons) to 2 (all the questions within the group of questions were answered correctly according to both scoring persons). Spearman correlation coefficient between the two scoring persons was high, 0.83, and most of the answers were rated identically by the two persons (79.5%).

The questionnaires were anonymous.

5.5 Data analysis

Answers of the students that filled both questionnaires were analysed. The biographical data (age and computer literacy) and students' self-reported attitude toward the seminar were analysed directly. The students' answers from the immediate and the retention knowledge tests were first scored and then analysed.

Differences between the traditional and the game groups, between boys and girls and between the immediate and the retention knowledge tests were tested separately for the G1- and the G2-questions using the mixed effects ANOVA. For the G2-questions we also used the two-way ANOVA for comparison of the groups and genders separately in the immediate test and in the retention test (Quinn & Keough, 2002). The effect size was expressed by Cohen's d and it was classified into negligible (Cohen's $d < 0.2$), small (Cohen's $d < 0.5$), medium (Cohen's $d < 0.8$) and large (Cohen's $d \geq 0.8$) (Cohen 1988).

Non-parametric statistics were used for evaluating the attitude toward the seminar, as these data were grossly non-normal. Wilcoxon rank sum test was used for testing differences between the groups (Quinn & Keough, 2002; Anděl, 2007). If needed, Bonferroni correction for multiple comparisons was used to ensure 0.05 level of significance (Weisstein, 1999-2011). Cliff's delta was used for estimating the effect size of the difference. The effect size was classified into four categories: negligible (Cliff's delta < 0.147), small (Cliff's delta < 0.33), medium (Cliff's delta < 0.474) and large (Cliff's delta ≥ 0.474). This classification corresponds to the commonly used classification of Cohen's *d* which represents the effect size for normally distributed data (Romano et al. 2006)

Two-sample test for equality of proportions ("prop.test" in statistical software "R") was used for comparison of percentages of students expressing positive attitude towards the seminar or the game.

6. Results

6.1 Comparison of the two groups with regard to biographical variables

We balanced the possible effects of class/school and gender across the two groups. Table 1 shows that the distribution of students from each class in the two groups was approximately equal as well as the distribution of gender.

Age and computer literacy of the students were obtained from the immediate post-questionnaire. Most of the students were either 15 or 16 years old (81%). None of them was younger than 15 and older than 18. Table 2 shows that both groups were similar in the age and in the biographical variables assessing their computer literacy.

--- Insert Table 2 about here ---

6.2 Comparison of the two groups with regard to the knowledge tests

There were two types of questions in the knowledge tests. The first type of questions (G1-questions) was aimed to test general knowledge of learning and memory. The second type of questions (G2-questions) was aimed to test knowledge about positive reinforcing. The knowledge addressed by the G2-questions was supposed to be strengthened by the microgame in the game group and by the additional lecture in the traditional group.

Results of the knowledge tests are presented in Figure 7 and in Tables 3, 4 and 5.

The scores from G1- and G2-questions were analysed separately by means of two mixed effects ANOVAs with "group" and "gender" as between subject factors and "test" as a within subject factor. The assumptions of normality and multi-sample sphericity were verified before performing the analysis.

The mixed effects ANOVA performed on G1-questions showed no difference between the groups ($F[1,96] = 1.4060$, $p = 0.2387$) and genders ($F[1,96] = 1.3190$, $p = 0.2536$) but difference between the tests ($F[1,96] = 111.5130$, $p < 0.0001$). All the interactions were non-significant (p's > 0.6542). Thus, the traditional and the game groups performed equally in the G1-questions although the effect size analysis showed small effect in favour of the traditional group in the immediate test (Table 3). Both groups reached higher scores in the immediate test than in the retention test with large effect size (Table 5).

With respect to G2-questions the mixed effects ANOVA showed significant difference between the groups ($F[1,96] = 6.4206$, $p = 0.0129$) but no differences between the genders ($F[1,96] = 1.0305$, $p = 0.3126$) and the tests ($F[1,96] = 1.0903$, $p = 0.2990$). All the interactions were non-significant (p's > 0.1428). The main finding of the second ANOVA is that the game group performed better than the traditional group in G2-questions. Additionally, the effect size analysis indicated that the difference was more pronounced in the retention test where the effect size was medium (Table 4) than in the immediate test where the effect size was small (Table 3). To further investigate this aspect, we separately analysed the scores from G2-questions in the immediate and in the retention tests using the two-way ANOVA with "group" and "gender" as between-subjects factors. In the immediate test no difference between the groups ($F[1,96] = 2.2661$, $p = 0.1355$) was found as well as no difference between the genders ($F[1,96] = 0.2581$, $p = 0.6126$) and no significant interaction ($F[1,96] = 1.7162$, $p = 0.1933$). On the contrary, the same analysis performed on the scores from G2-questions in the retention test showed highly significant difference between the groups ($F[1,96] = 11.0253$, $p = 0.0013$), but no difference between the genders ($F[1,96] = 2.1577$, $p = 0.1451$) and no interaction ($F[1,96] = 0.0083$, $p = 0.9278$).

Concerning the difference between the scores in the immediate and in the retention tests, it has been shown by the mixed effects ANOVA for G2-questions (presented above)that the factor "test" was non-significant, suggesting that the scores in the two testes were similar. However, even though the effect size analysis comparing the scores across the tests showed indeed only negligible effect for the game group, it also a showed small effect for the traditional group (Table 5). We will return to this point bellow and in the discussion.

In general, the mixed effects ANOVAs showed no difference between boys and girls. Nevertheless, the effect size analysis presented in Tables 3 and 5 indicated some possible differences, which may be confirmed or excluded if larger number of students was tested.

The effect size analysis comparing scores of the traditional and the game groups in G1-questions in the immediate test showed small effect for boys and only negligible effect for girls (Table 3). In this case Cohen's *d* was similar for both genders and it just failed to reach the limit for being classified as small in girls. In other words, the effect size analysis confirmed there is no difference between girls and boys in G1 questions in the immediate test.

Larger effect size discrepancy between genders was found in G2-questions in the immediate test. There was medium effect size when boys from the two groups were compared but negligible effect size when girls were compared (Table 3). Boys from both groups and girls from the game group obtained similar scores in the retention test as they did in the immediate test according to the effect size analysis (Table 5). On the contrary, girls from the traditional group obtained lower scores in the retention test in comparison to the immediate test. The effect size of this difference was small, however, it almost reached the limit for being classified as medium (Table 5). Thus, the data indicates that the small difference regarding effect size between G2-questions' scores in the immediate test and the retention test can be attributed to girls rather than boys.

--- Insert Figure 7 about here ---

--- Insert Table 3 about here ---

--- Insert Table 4 about here ---

--- Insert Table 5 about here ---

6.3 Comparison of the two groups with regard to appeal to students

The attitude of the students toward the whole seminar was assessed at the end of the seminar by two questions. The students scored the overall appeal towards the seminar and the subjective educational value of the seminar. In addition, the students from the game group were asked "How did they like the game". In all the questions the students choose one out of four ordinal levels ranging from 1 to 4. The two highest scores "3" and "4" indicated positive judgement.

The distribution of students' answers to the overall appeal toward the seminar is shown in Figure 8. The distributions of the two groups are similar. Most of the students responded by marking the level 3. Medians, means and standard deviations as well as group comparison is presented in Table 6. There was no difference between the groups as measured by Wilcoxon rank sum test (Table 6) and the effect size expressed by Cliff's delta was negligible (Table 6). The results were similar if only boys or only girls were evaluated (Table 6). We want to stress that vast majority of the students regardless of whether they were from the traditional or from the game group answered by score 3 or 4 (Table 7). Thus their judgement about the overall appeal toward the seminar was positive.

The subjective educational value of the whole seminar is presented in Figure 8 and in Table 6. The traditional group scored higher regardless of whether boys, girls or both genders together were evaluated (Table 6). This difference almost reached significance when both genders were analysed together. If Bonferroni correction for multiple comparisons (6 tests in this case) was not used then the difference was highly significant (Table 6). However, the effect size of this difference was small (Table 6). The percentage of students who expressed positive judgement about the educational value of the seminar (students who answered by score 3 or 4) is shown in Table 7. The judgement was positive in most of the students regardless of the group. Note that lower percentage of girls in the game group expressed positive judgement about the educational value of the seminar in comparison to boys from the same group as well as in comparison to girls from the traditional group (Table 7). If the percentage of girls expressing positive judgement in the game group was compared with the percentage of girls expressing positive judgement in the traditional group by two-sample test for equality of proportions then the difference was significant, but only without correction for multiple testing (chi-squared = 4.578, df = 1, p = 0.0324).

Table 8 presents percentages of students from the game group who expressed positive judgement toward the game. It is important that most of the students liked the game regardless of the gender although the percentage was slightly lower for girls than for boys (Table 8).

--- Insert Figure 8 about here ---

--- Insert Table 6 about here ---

--- Insert Table 7 about here ---

--- Insert Table 8 about here ---

7. Discussion

This pilot study evaluated the usage of an educational biology micro-game Orbis Pictus Bestialis employed as a supplement after a traditional biology lecture at four high-school classes in the Czech Republic. The topic of the game and of the lecture, animal learning, is an integral part of the Czech national high-school curricula. The purpose of the game was to reinforce and integrate part of the knowledge learnt in the introductory lecture. The game was compared against an extra lecture that used media-rich materials. Both the game and the extra lecture were followed by a short debriefing. The teacher guided the students during playing the game in the game group. The learning in OPB is integral part of game-playing; the game is not a "drill-and-practice" software. We have intentionally chosen this specific type of the game's usage because broader literature on games and simulations, including non-computer ones, suggests that this format can serve well for educational purposes as discussed in Sec. 2.

7.1 Major findings

Based on the review of literature, we investigated four hypotheses:

1) Our game will be at least as effective as traditional instruction concerning the gain of knowledge that the game is supposed to reinforce and integrate, if measured immediately after the exposure.

This hypothesis is confirmed; there is a small positive, though not significant, effect of the game on the questions that tested this knowledge (Table 3, Figure 7). This result is not very surprising as it is in line with many previous findings (see, e.g., a recent review of Hays, 2005; an older review on simulation gaming of Bredmeier & Greenblat, 1981; and recent studies, e.g., Wrzesien et al., 2010; Huizenga et al., 2009; Annetta et al.., 2009). Actually, the positive effect seems to be more pronounced for boys, but there seems to be no effect for girls, a point to which we will return in Section 7.2.

2) The reinforced and integrated knowledge will be retained better by the game group.

This hypothesis is also confirmed; there is a medium positive effect size of the game on the questions that tested this knowledge a month after the treatment, regardless of whether girls, boy or both genders together were analysed (Table 4, Figure 7). The difference between the game and the traditional group is significant. This is the key finding of the present study as the data on retention are relatively limited. Our results are in line with reports of a few past studies on simulation gaming that reported retention data (see, e.g., Pierfy, 1977) and they also agree with results of the recent study of an educational game on physiological concepts (Wong et al., 2007) and a study of a historical strategy COTS game (Egenfeldt-Nielsen, 2005). Also Venneman & Knowles (2005) reported that the Sniffy simulation had positive effects on students' knowledge, as measured after the school term, and Huizenga et al. (2009, and also personal communication) reported positive effect of their game using one-week post-tests but that study employed shorter treatment for the control group than for the experimental group.

In general, this result is more important than the first one because long-term effects of curricular education are more crucial than short-term effects.

An open question is when should be the long-term effects measured, most studies mentioned above differ in time of the follow-up administration, ranging from one-week after the treatment (Wong et al.) to five months (Egenfeldt-Nielsen).

3) The game will not be detrimental to knowledge that was transmitted by the expository formal lecture but not reinforced and integrated by the game; neither immediately nor later.

This hypothesis is also confirmed. In general, subjects from the game group scored worse on the questions that tested this knowledge, both immediately and a month after the treatment, than subjects from the traditional group, but the differences between the groups, are negligible or small and not significant (Table 3, 4; Figure 7). This result is interesting for two reasons. First, we employed a specific type of the game's usage in the formal schooling system. Should this type become one of standard formats of usage in this context, it is important to know that the game is not detrimental to information from the lecture that was not subsequently strengthened by the game. Second, to our knowledge, this question has never been tested explicitly, though results of some past studies can be re-analysed regarding it. For instance, Ioannidou et al. (2010) reported large or very large positive effects of a game on cardiopulmonary system on a subset of questions that were arguably mostly related to the game activities, as opposed to the rest of the questions, on which only negligible positive effect was found. When re-analysed in detail, such kind of data may further support our Hypothesis 3. Note, however, that Ioannidou et al. administered post-tests "just after the [lesson] unit" (p. 158).

4) The course with the game will be more engaging than the lecture-only course.

This hypothesis is not confirmed. There is no effect of the game with regard to the overall appeal of the whole course and there is a small negative effect of the game concerning the perceived educational value of the game and this is true for both genders. This effect approaches significance (Table 6). This result is mildly surprising since many previous studies reported that learners are more interested in simulation/game activities than in conventional classroom instruction (see, e.g., Hays, 2005 for a review, and Annetta et al., 2009 for a recent game study). However, for instance the review of Lee (1999) reported five studies in which students showed negative or very little preference to simulations. Also some recent studies reported mixed results concerning affective variables (e.g., Huizenga et al., 2009). The results of different studies may differ due to several reasons; most notably due to slightly different questions on affective variables, different type of control (we used a power-point presentation with colour figures and a video, a point to which we will return below) and different age of learners. As already said, adolescents are more difficult to engage in school learning and harder to motivate than younger children (e.g. Eccles & Midgley, 1989; cited from Papastergiou, 2009). They may also perceive games as a leisure time activity with limited pedagogical value (e.g. de Freitas, 2006), which might not be the case of younger students. However, the most important point is that the vast majority of the students from the game group expressed positive attitude toward the whole seminar (92%) and toward the game (90%) and considered the educational value of the seminar positively (85%) (see Sec. Comparison of the two groups with regard to appeal to students).

7.2 General comments on possible gender differences

Learning effects of educational games and simulations have only rarely been investigated with respect to gender. The likely reason is that many studies produced unequivocal results and their analysis regarding gender might confuse the situation further.

However, the issue of gender differences is important because it is known that women in general tend to prefer different game genres than men and tend to have different motivations than men for playing computer games (e.g., van Looy & Courtois, 2010; Hartmann & Klimmt, 2006) and different impact of educational games on boys and girls can thus be expected. For instance, Egenfeldt-Nielsen (2005) reported that high-school girls tended to have larger problems than boys with mastering interface of a complex COTS game, on the other hand, Papastergiou (2009) in a study of a more “quiz-based” game found the game to be equally motivational for both girls and boys and that learning gains also did not differed significantly between genders.

Thus, we have analysed our data also with respect to gender. We found no noticeable difference between genders regarding cognitive variables, except for one fact. The game was more effective for reinforcing and integrating knowledge on positive reinforcement for boys than for girls when measured immediately after the treatment (Table 3). In fact, the effect size analysis suggests no difference between girls from the traditional group and the game group in that test (as opposed to boys). However, this gender difference diminishes in the retention test, where both genders show medium size gain in the game group comparing to the traditional group (Table 4).

Concerning affective variables, the feedback from both genders was highly positive in both groups (Table 6). Girls in general were slightly more critical towards to game than boys but this difference was not significant when corrected for multiple comparisons.

One possible explanation of the small gender differences is that the difference in immediate tests between girls and boys might correspond to slightly lower appeal of the game for girls; however, we do not want to make strong claims regarding this point as our study was not focused on investigation of gender differences. More research is needed to investigate this important issue. Our key result is that there is a positive medium size significant difference between the two groups regarding knowledge the game is supposed to reinforce and integrate in retention tests and this result is robust.

7.3 Comments on the control (traditional) group

In educational game studies, type of control is a variable that is only rarely studied systematically. The effect of a game may differ when comparing to, for instance, a chalk-and-blackboard formal lecture or a self-study using media-rich materials. In this regard, the study of Wong et al. (2007) reported useful information. They evaluated their game on physiology concepts for undergraduate college students employing "2 x 2 between subject follow-up design, with the factors of interactive and media richness" (p. 51). Essentially, they compared a game (interactive, media-rich) against a) a video of a replay of the game (non-interactive, high media-richness), b) a "hypertext version of the game including narrative context and screenshots from the game" (interactive, moderate media-richness), and c) "educational text fragments from the game on the computer screen allowing subjects to read at their own pace and in the order of their choice" (non-interactive, low-media richness) (p. 51). That study reported significant positive effect of the game on knowledge gains as well as on several affective variables, but only against the text-only condition. There were a) only negligible differences against the hypertext version (interactive, moderate media-richness) in knowledge gains and in the affective domain, and b) only negligible differences against the replay version (non-interactive, high media-richness) in knowledge gains and small positive but often not significant effect of the game regarding the affective variables. The knowledge gain results sustained over a week (the affective variables were measured only once). Note that the information about debriefing and expository lecture was not given in that paper.

Our control group received media-rich treatment in a passive way; this resembles Wong et al.'s replay group. Thus, concerning immediate knowledge gains and engagement, our results are more or less in line with theirs. However, Wong et al. found no differences between the game group (interactive) and the replay group (non-interactive) in knowledge gains even regarding retention, which is intriguing. Our findings differ from their in this regard. One possible reason is that they administered one-week post-test whereas we employed one-month follow-up. Perhaps it is better to administer post-tests after a longer period? Note that other studies suggesting that interactive treatment has better outcome regarding knowledge gain than non-interactive one also exist (see, e.g., Kinzie et al., 1993, who found that students interfacing with an interactive simulation of a frog dissection performed a subsequent dissection more effectively than students viewing a videotape).

More research is needed to elucidate this point. This is important because developing interactive hypertexts and non-interactive videos is often cheaper than development of games.

7.4 Limitations

The present study has several limitations. One limitation concerns the measurement instruments. Given the objective of the game, perhaps the best way to assess the students' knowledge would be to test transfer: whether the students are able to train real animals (see e.g. Annette et al., 2009, p. 79; Reif, 2008, ch. 21.6). However, that would be very demanding on the amount of naive untrained animals, their transportation to the schools, etc. Additionally, we were also interested in the effect of the playing on the factual knowledge gained during the formal lecture but not strengthened by the game. Thus, we employed classical paper-and-pencil questionnaires. For future work, transfer tests would be an advantage. These tests may not necessarily employ real animals, but can be performed, for instance, using other game levels.

It would have been also an advantage to have information about students' attitudes towards biology and animals and previous knowledge on the topic of animal learning. Unfortunately, the information about students' general grades was not available to us. Additionally, students' knowledge on animal learning is known to be low in general (in the Czech high-schools) and no marks on exactly this topic are available. We had to use post-tests only design to avoid practise effect and cuing of students on what should be remembered (see, e.g., Judd et al., 1991). However, we could have asked for information about whether a student has a pet and any direct experience with its training. In fact, we asked for this information during the lectures verbally and our informal observation is that (a few) students having pets can be divided into two groups; “enthusiasts” who are very keen to study this topic, and “bored” who pompously demonstrates that they know the material better than the teacher and thus need not pay attention at all. Therefore, more information about students’ personality types, attitude towards the topic, and what experience they already have with training a pet would have helped to elucidate conditions under which the game is effective. Other way to improve the design can be administering knowledge-based pre-test several weeks before the treatment; however, this might be problematic for time-constraints of the high-schools.

The present study also has several confounding and uncontrolled variables, for instance the time of the day in which the treatment was administered or what has happened between the treatment and the follow-up. Such variables are common in quasi-experimental field studies of this kind. One possibility to reduce the number of confounding variables is to use a controlled laboratory environment, in which the treatment is administered to exactly one subject at a time, similarly to many psychological experiments. This, however, is less natural than the classroom environment, a trade-off, and does not mitigate the problems with the period between the treatment and the follow-up.

Finally, our study produced quantitative outcomes, a kind of aggregate descriptions of *what* happened during the treatments. Such outcomes can be complemented with more qualitative data that would help to elucidate *how* our outcome was achieved. Both quantitative and qualitative approaches are needed; they are complementary rather than competitive. The approach chosen depends on the experimental question and because only a few controlled studies exist regarding retention of materials learnt using educational computer games, we have chosen the quantitative approach. However, in future, we aim at supplementing our research framework with focus groups. Additionally, in our opinion, qualitative research can be particularly helpful for elucidating possible gender differences.

None of these limitations undermines the results presented. Rather, they open ways for future work.

7.5 Possible generalization

An important question is to which extent our results can be generalized. It is known that results of a single study on usage of an educational game cannot be generalized easily to other games used differently and in a different context (see, e.g., Hays, 2005, for more on this point). This is not a problem posed by design of the studies but by the complexity of games.

One issue is the target group. In our opinion, results of the present study can be only generalized to those micro-games that are used at high-schools, and perhaps at the undergraduate level. Another issue is what the game teaches. The present study and the game focus on teaching the core knowledge of a particular topic. Thus, the results cannot be generalized straightforwardly for teaching high-level skills such as scientific problem solving (see., e.g., McFarlane & Sakelariou, 2002; Barab et al., 2007; Squire & Jan, 2007). Yet another issue is the format of usage of the game; we can generalize the results only to micro-games that are used similarly to our game. For instance, it is known that the results may differ based on the level of guidance or based on when the game (or simulation) is administered during the course (e.g. Rivers & Vockel, 1987; Brant et al., 1991; Lee, 1999; Tobias &

Fletcher, 2007). It is also known that a good debriefing plays an important role (e.g. Peters & Vissers, 2004; Hays, 2005). Still another issue is the type of the game. We investigated a micro-game, in which, by definition, the learning is integral part of the playing (and vice versa). Our results can be most likely generalized for interactive single-user simulations, but caution is important when generalizing to complex games with days or weeks long game-play (e.g., Buch & Egenfeldt-Nielsen, 2006; Brom et al., 2010), or to drill-and-practise edutainment (e.g., Papastergiou, 2009, but see also Seelhammer & Niegemann, 2009). Larger games require completely different allocation of learners' attention and drill-and-practise software tend to rely on extrinsic motivation while micro-games capitalize more on intrinsic motivation.

Thus, one of the first next steps is to replicate our findings using the same usage format but a different intrinsically motivating software, perhaps a Netlogo simulation (cf. Wilensky & Novak, 2010) or another micro-game. Next steps also include systematic varying of variables determining the usage format of the software. It may be perhaps uncomfortable for some serious game proponents that the issue of "usefulness of serious games" can not be reconciled by a single study, but this issue is basically more complicated than many people would wish.

8. Conclusion

This pilot study investigated one possible approach to harnessing educational computer games at high-schools for the purpose of teaching core knowledge of a particular topic. This approach capitalizes on using a micro-game that can be played with a teacher's guidance as a supplement after a traditional lecture. The game's purpose is to reinforce and integrate part of the knowledge learnt in the expository lecture. The game is supposed to be followed by a short debriefing.

To evaluate feasibility of this approach, we employed a game on the topic of animal learning in the present study. The game-group was compared to a group in which the expository lecture was followed by an extra lecture that used media-rich materials, including one video. The time allotment was the same in both groups. Together, we analysed data from 100 students from 3 different high-schools in the Czech Republic.

The results showed that the game-playing is comparable to the more traditional form of teaching concerning knowledge gains when measured immediately after the treatment and has a significant medium positive effect size regarding retention as measured one month after the treatment. The results also suggested that the game-playing is not detrimental to information transmitted in the expository lecture but not strengthened by the game. Finally, the whole seminar with the game was found appealing by the students and with enough educational value, even though the perceived educational value was significantly lower than in the control group with small effect size.

Our results also suggested that gender differences - for our type of game and usage format - are rather small; the perceived educational value was slightly lower for girls than for boys and there was medium positive immediate effect size of the game on boys but not on girls. However, these effects were not significant. The main outcomes apply for both genders.

Taken together, the data indicate that the proposed format of harnessing educational computer games at high-schools is promising. The next steps include replicating the findings with another simulation and varying variables regarding the usage format.

References

Aitkin, A. L. (2004) *Playing at Reality: Exploring the Potential of the Digital Game as a*

Medium for Science Communication, PhD thesis, Australian National University.

Akpan, J. P. (2001) 'Issues Associated with Inserting Computer Simulations into Biology Instruction: A Review of the Literature', *Electronic Journal of Science Education*, 5(3), 1-33.

Alloway, T., Wilson, G. and Graham, J. (2005) *Sniffy The Virtual Rat*, Belmont: Wadsworth/Thomson Learning.

Anděl, J. (2007) *Statistical methods*, 4th ed., MatfyzPress (in Czech)

Annetta, L. A., Minogue, J., Holmes, S. Y. and Cheng, M. (2009) 'Investigating the impact of video games on high school students' engagement and learning about genetics', *Computers & Education*, 53(1), 74-85.

Arnseth, H. C. (2006) 'Learning to Play or Playing to Learn - A Critical Account of the Models of Communication Informing Educational Research on Computer Gameplay', *Game Studies*, 6(1), available: http://www.gamestudies.org/0601/articles/arnseth [accessed 17.11.2010].

Barab, S., Thomas, M., Dodge, T., Carteaux, R. and Tuzun, H. (2005) 'Making learning fun: Quest Atlantis, a game without guns', *Educational Technology Research & Development*, 53(1), 86-107.

Barab. S., Sadler, T. D., Heiselt, C., Hickey, D. and Zuicker S. (2007) 'Relating Narrative, Inquiry, and Inscriptions: Supporting Consequential Play', *Journal of Science Education and Technology*, 16(1), 59-82.

Bear, M. F., Connors, B. W. and Paradiso, M. A. (2007) *Neuroscience: Exploring the Brain,* Lippincott: Williams & Wilkins.

Becker, K. (2008) *The Invention of Good Games: Understanding Learning Design in Commercial Video Games*, PhD thesis, University of Calgary.

Blackham, D. (1974) *Operant Conditioning: An Experimental Analysis of Behavior,* London: Methuen & Ltd.

Bloom, B. S., ed. (1956) *Taxonomy of Educational Objectives, Handbook 1: Cognitive Domain,* New York: David McKay.

Brant, G., Hooper, E. and Sugrue, B. (1991) 'Which comes first: The simulation or the lecture? ', *Journal of Educational Computing Research,* 7(4), 469-481.

Bredemeier, M. E. and Greenblat, C. S. (1981) 'The educational effectiveness of simulation games: A synthesis of findings.', *Simulation & Games*, 12(3), 307-332.

Brom, C., Sisler, V. and Slavík, R. (2010) 'Implementing Digital Game-Based Learning in Schools: Augmented Learning Environment of 'Europe 2045' ', *Multimedia Systems,* 16(1), 23-41.

Buch, T. and Egenfeldt-Nielsen, S. (2006) 'The Learning Effects of Global Conflicts' in *Conference Proceedings Media@Terra*, Athens.

Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences (second ed.). Lawrence Erlbaum Associates.

Cordova, D. I. and Lepper, M. R. (1996) 'Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice', *Journal of Educational Psychology,* 88, 715-730

Dorn, D. S. (1989) 'Simulation games: One more tool on the pedagogical shelf', *Teaching Sociology,* 17(1), 1-18.

Eccles, J. and Midgley, C. (1989) 'Stage-environment fit: Developmentally appropriate classrooms for young adolescents' in Ames, C. and Ames, R., eds., *Research on motivation in education* New York: Academic Press, 139 - 186.

Egenfeldt-Nielsen, S. (2005) *Beyond Edutainment: Exploring the Educational Potential of Computer Games*, PhD thesis, University of Copenhagen.

Facer, K., Ulicsak, M. and Sandford, R. (2007) 'Can Computer Games Go to School ...?', in *Emerging Technologies for Learning* 2, 47 - 63, available: http://partners.becta.org.uk/page_documents/research/emerging_technologies07.pdf [accessed 16.11. 2010].

Faria, A. J. (2001) 'The changing nature of business simulation/gaming research: A brief history', *Simulation & Gaming* 32(1), 97-110.

de Freitas, S. (2006) Learning in immersive worlds. *Joint information systems committee*, available: http://www.jisc.ac.uk/eli_outcomes.html [accessed 16.11. 2010].

Gee, J. P. (2003) *What video games have to teach us about learning and literacy,* New York: Palgrave/St. Martin’s.

Hays, R. T. (2005) *The Effectiveness of Instructional Games: A Literature Review and Discussion*, *Technical Report 2005-004*, Orlando: Naval Air Warfare Center Training Systems Division.

Huizenga, J., Admiraal, W., Akkerman, S. and ten Dam, G. (2009) 'Mobile game-based learning in secondary education: engagement, motivation and learning in a mobile-city game', *Journal of Computer Assisted Learning,* 25(4), 332-344.

Church, R. and Hand, S. (1992) 'Industrial melanism in moths: An interactive computer simulation to demonstrate evolutionary theory', *Journal of Biological Education,* 26(1), 34-36.

Ioannidou, A., Repenning, A., David Webb, D. K., Luhn, L. and Daetwyler, C. (2010) 'Mr. Vetro: A Collective Simulation for teaching health science ', *International Journal of Computer-Supported Collaborative Learning*, 5(2), 141-166.

Judd, C. M., Smith, E. R. and Kidder, L. H. (1991) *Research Methods in Social Relations,* Fort Worth: Hartcourt Brace.

Ketelhut, D. J., Schifter, C. C. (2011) 'Teachers and game-based learning: Improving understanding of how to increase efficacy of adoption', *Computers & Education*, 56, 539-546.

Kinzie, M. B., Strauss, R. and Foss, J. (1993) 'The effects of an interactive dissection simulation on the performance and achievement of high school biology students', *Journal of Research in Science Teaching,* 30(8), 989-1000.

Kirriemuir, J. and McFarlane, A. (2004) 'Literature Review in Games and Learning', Nesta Futurelab series, Report 8, Bristol: Futurelab.

Kirschner, P. A., Sweller, J. and Clark, R. E. (2006) 'Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching', *Educational Psychologist,* 41(2), 75 - 86.

Klopfer, E. (2008) *Augmented Learning: Research and Design of Mobile Educational Games,* Cambridge: The MIT Press.

Lee, J. (1999) 'Effectiveness of computer-based instructional simulation: a meta analysis', *International Journal of Instructional Media,* 26(1), 71-85.

Lutterschmidt, W. I. and Schaefer, J. F. (1997) 'A computer simulation for demonstrating and modelling predator-prey oscillations', *Journal of Biological Education,* 31(3), 221-227.

Malone, T. W. (1981) 'Toward a theory of intrinsically motivating instruction', *Cognitive Science*, 5(4), 333-369.

McFarlane, A. and Sakellariou, S. (2002) 'The Role of ICT in Science Education', *Cambridge Journal of Education*, 32(2), 219-232.

Mikropoulos, T. A., Katsikis, A., Nikolou, E. and Tsakalis, P. (2003) 'Virtual environments in biology teaching', *Journal of Biological Education,* 37(4), 176-181.

Niemiec, R. and Walberg, H. J. (1985) 'Computers and Achievement in the Elementary

School', *Journal of Educational Computing Research*, 1, 435-440.

Nilsson, E. M. and Svingby, G. (2009) 'Gaming as Actions: Students Playing a Mobile Educational Computer Game', *HUMAN IT*, 10(1), 26-59.

Papastergiou, M. (2009) 'Digital Game-Based Learning in high school Computer Science education: Impact on educational effectiveness and student motivation', *Computers & Education,* 52, 1-12.

Pavlov's Dog, an educational simulation (2001) *Nobelprize.org*, available: http://nobelprize.org/educational_games/medicine/pavlov [accessed 16.11. 2010].

Peters, V. A. M. and Vissers, G. A. N. (2004) 'A simple classification model for debriefing simulation games', *Simulation & Gaming*, 35(1), 70-84.

Pierfy, D. A. (1997) 'Comparative simulation game research: Stumbling blocks and stepping stones', *Simulation and Games,* 8(2), 255-268.

Quinn, G. P. and Keough, M. J. (2002) *Experimental Design and Data Analysis for Biologists*, 1st ed., Cambridge University Press.

Randel, J. M., Morris, B. A., Wetzel, C. D. and Whitehill, B. V. (1992) 'The effectiveness of games for educational purposes: A review of recent research', *Simulation & Gaming,* 23(3), 261-276.

Reif, F. (2008) *Applying Cognitive Science to Education. Thinking and Learning in Scientific and Other Complex Domains,* Cambridge: The MIT Press.

Rieber, L. P. (1996) 'Seriously considering play: Designing interactive learning environments based on the blending of microworlds, simulations, and games', *Educational Technology, Research, and Development*, 44(2), 43-58.

Rivers, R.-H. and Vockell, E. (1987) 'Computer simulations to stimulate scientific problem solving', *Journal of Research in Science Teaching*, 24(5), 403-416.

Romano, J., Kromrey, J. D., Coraggio, J. and Skowronek, J. (2006) 'Appropriate statistics for ordinal level data: Should we really be using t-test and Cohen's d for evaluating group differences on the NNSE and other surveys?', in *Annunal meeting of the Florida Association of Institutional Research*, Florida.

Rosenbaum, E., Klopfer, E. and Perry, J. (2007) 'On Location Learning: Authentic applied Science with Networked Augmented Realities', *Journal of Science Education and Technology*, 16(1), 31-45.

Sandford, R., Ulicsak, M., Facer, K., Rudd, T. (2006) *Teaching with Games: Using commercial off-the-shelf computer games in formal education*, Bristol: Futurelab, available: http://www.futurelab.org.uk/resources/documents/project_reports/teaching_with_games/TWG_report.pdf [accessed 16.11. 2010].

Seelhammer, C. and Niegemann, H. M. (2009) *Playing Games to Learn – Does it actually work?*, translated by Kong, S. C., et al., Hong Kong: Asia-Pacific Society for Computers in Education

Sørensen, B. H., Danielsen, O. and Nielsen, J. (2007) 'Children's informal learning in the context of schools of the knowledge society', *Education and Information Technologies*, 12(1), 17-27.

Squire, K. (2004) *Replaying history: Learning World History through playing Civilization III*, PhD thesis, Indiana University.

Squire, K. (2005) *Game-based Learning: Present and Future State of the Field. An x-Learn Perspective Paper*, MASIE Center, available: https://files.pbworks.com/download/X5sRuw315P/immersionfall09/12947894/Squire--2005--Game-Based_Learning.pdf [accessed 16.11. 2010].

Squire, K. and Jan, M. (2007) 'Mad City Mystery: Developing Scientific Argumentation Skills with a Place-based Augmented Reality Game on Handheld Computers', *Journal of Science Education and Technology*, 16(1), 5-29.

Šisler, V. and Brom, C. (2008) 'Designing Educational Game: Case Study of 'Europe 2045'' in *Transactions on Edutainment I*, New York: Springer-Verlag Berlin Heidelberg, 1-16.

Thomas, R. and Hooper, E. (1991) 'Simulations: An opportunity we are missing', *Journal of Research on Computing in Education*, 23(4), 497 – 513.

Tobias, S. and Fletcher, J. D. (2007) 'What Research Has to Say about Designing Computer Games for Learning', *Educational Technology Magazine: The Magazine for Managers of Change in Education*, 47(5), 20-29.

VanSickle, R. L. (1986) 'A quantitative review of research on instructional simulation gaming: A twenty-year perspective', *Theory and Research in Social Education*, 14(3), 245-264.

Venneman, S. S. and Knowles, L. (2006) 'Sniffing Out Efficacy: Sniffy Lite, a Virtual Animal Lab', *Teaching of Psychology*, 32(1), 66-68.

Watkins, C. J. C. H. (1989) *Learning from Delayed Rewards*, PhD thesis, Cambridge University.

Weisstein, E. W. (1999 - 2011) 'Bonferroni Correction', *MathWorld: A Wolfram Web Resource*, available: http://mathworld.wolfram.com/BonferroniCorrection.html [accessed 4.3.2001].

Wilensky, U. (1999) NetLogo, Center for Connected Learning and Computer-Based Modeling, Northwestern University, availible: http://ccl.northwestern.edu/netlogo/ [accessed 16.11. 2010].

Wilensky, U. and Novak, M. (2010) 'Understanding evolution as an emergent process: learning with agent-based models of evolutionary dynamics' in Taylor, R. S. and Ferrari, M., eds., *Epistemology and Science Education: Understanding the Evolution vs. Intelligent Design Controversy*, New York: Routledge.

Wong, W. L., Shen, C., Nocera, L., Carriazo, E., Tang, F., Bugga, S., Narayanan, H., Wang, H. and Ritterfeld, U. (2007) *Serious video game effectiveness*, translated by Inakage, M., Lee, N., Tscheligi, M., Bernhaupt, R. and Natkin, S., Salzburg: 49-55.

Wrzesien, M. and Raya, M. A. (2010) 'Learning in serious virtual worlds: Evaluation of learning effectiveness and appeal to students in the E-Junior project', *Computers &*

Education, 55(1), 178-187.

Figures

A note: all the figures can be printed in black & white.

Figure 1. A screenshot from the game.

Figure 2. Associating the sound of the clicker (in the upper left corner) with the food stimulus.

Figure 3. The dog has learned that the contact between its paw and the trainer's hand is rewarded. It raises its paw in order to make the contact.

Figure 4. The dog attempts to touch the trainer's hand positioned high above by raising its paw. When the paw is sufficiently high according to the player's demand, the dog is rewarded without making the contact with the hand. If the reward is delayed then the dog starts to jump in order to reach the hand.

Figure 5. Pronouncing the verbal command to be associated with the visual stimulus (the hand).

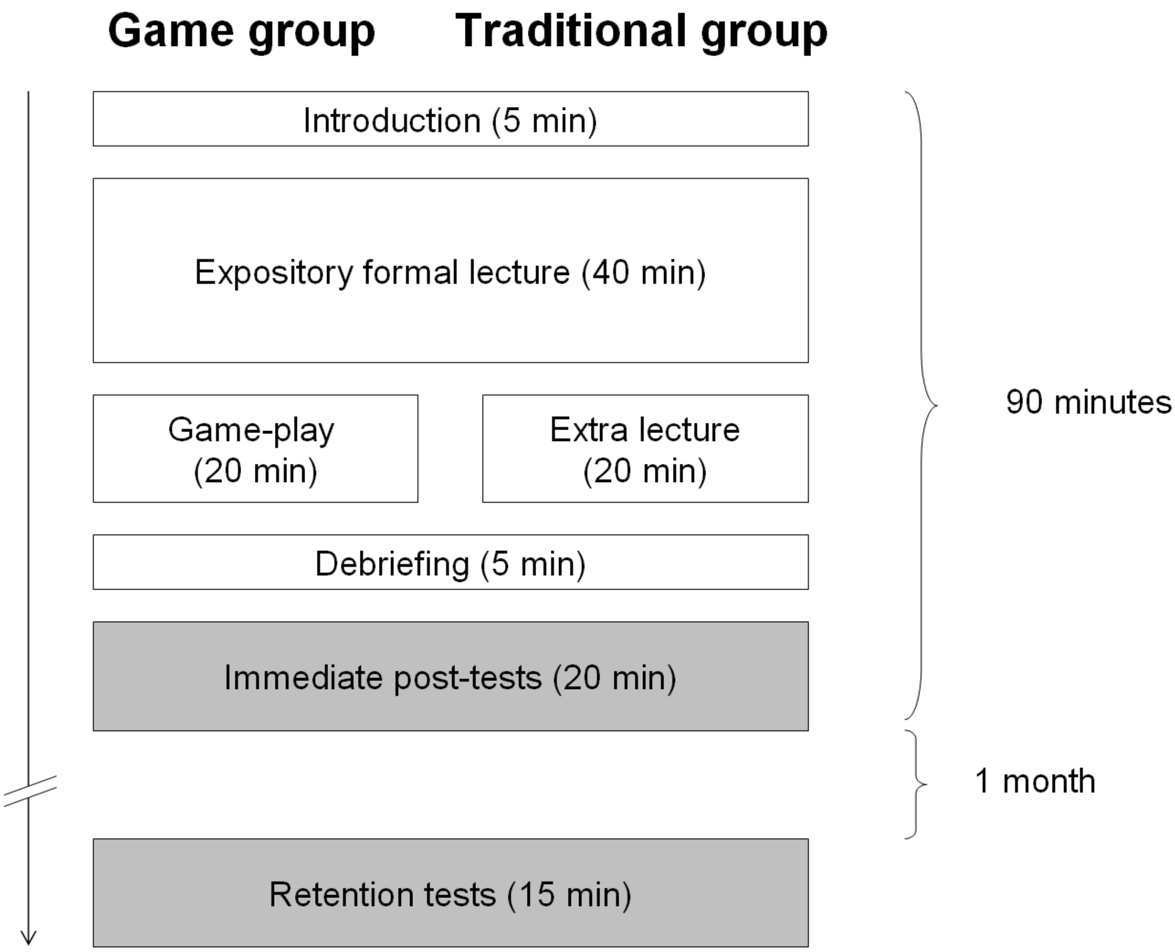

Figure 6. Experimental design.

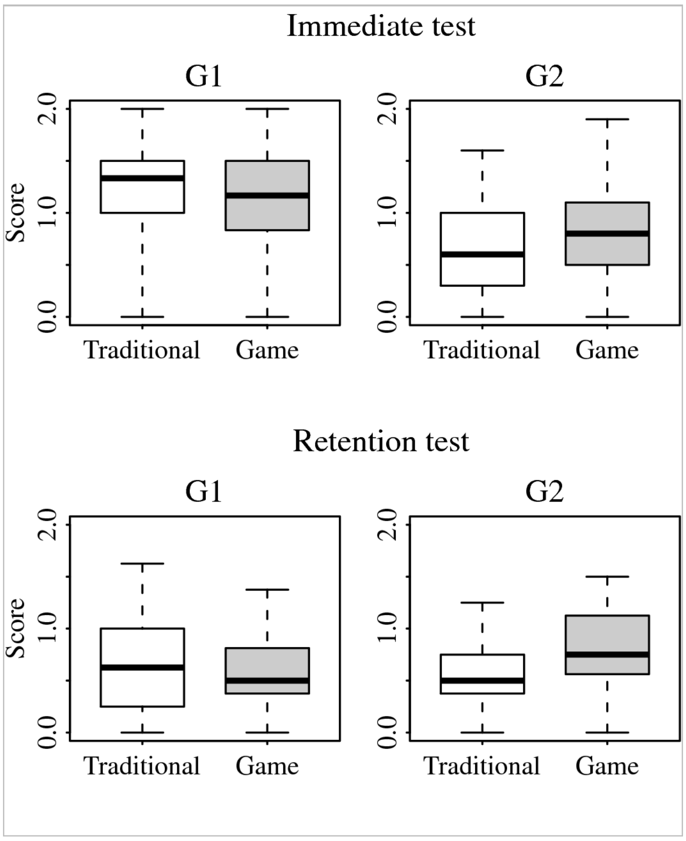

Figure 7. Performance of the traditional and the game groups in the knowledge tests. Medians, 1st and 3rd quartiles and ranges are shown.

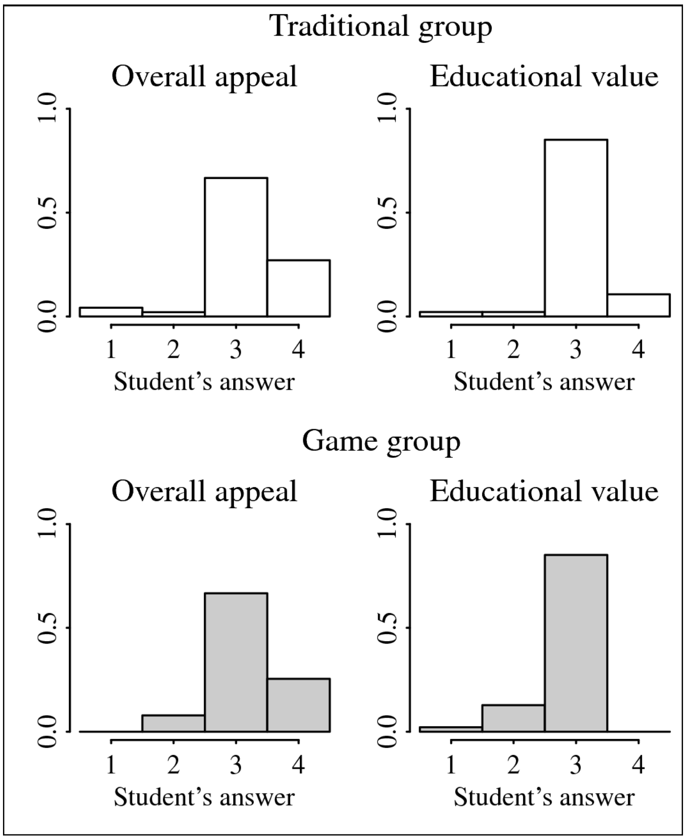

Figure 8: Distributions of students' answers in the questions assessing their overall appeal and subjective educational value of the lecture.

Figure captions separately

Figure 1. A screenshot from the game.

Figure 2. Associating the sound of the clicker (in the upper left corner) with the food stimulus.

Figure 3. The dog has learned that the contact between its paw and the trainer's hand is rewarded. It raises its paw in order to make the contact.

Figure 4. The dog attempts to touch the trainer's hand positioned high above by raising its paw. When the paw is sufficiently high according to the player's demand, the dog is rewarded without making the contact with the hand. If the reward is delayed then the dog starts to jump in order to reach the hand.

Figure 5. Pronouncing the verbal command to be associated with the visual stimulus (the hand).

Figure 6. Experimental design.

Figure 7. Performance of the traditional and the game groups in the knowledge tests. Medians, 1st and 3rd quartiles and ranges are shown.

Figure 8: Distributions of students’ answers in the questions assessing their overall appeal and subjective educational value of the lecture.

Tables

		Traditional group	Game group	Total
School A, Class 1	Boys	9	8	17
	Girls	5	6	11
	Total	14	14	28
School A, Class 2	Boys	7	8	15
	Girls	6	5	11
	Total	13	13	26
School B, Class 3	Boys	9	7	16
	Girls	4	5	9
	Total	13	12	25
School C, Class 4	Boys	5	6	11
	Girls	4	6	10
	Total	9	12	21
All students	Boys	30	29	59
	Girls	19	22	41
	Total	49	51	100

Table 1: Distribution of participants (of which the data were analysed) according to gender, school/class and intervention group

		Age	Frequency of computer use	Frequency of game use
Traditional group	Median	16	4	2
	Mean	16.1	3.33	2.31
	SD	1.01	0.77	0.98
Game group	Median	16	3	2
	Mean	15.82	3.31	2.4
	SD	0.68	0.68	1.12
Effect size:	Cliff's delta	0.12	0.03	-0.04
	effect	negligible	negligible	negligible

Table 2: Comparison of the two groups as to biographical variables

		G1			G2		
		Boys	Girls	All	Boys	Girls	All
Traditional group	Mean	1.27	1.16	1.23	0.62	0.7	0.65
	SD	0.52	0.55	0.53	0.48	0.48	0.48
Game group	Mean	1.15	1.05	1.11	0.86	0.7	0.79
	SD	0.51	0.57	0.53	0.43	0.46	0.45
Effect size	Cohen's d	-0.24	-0.19	-0.23	0.54	-0.01	0.3
	effect	small	negligible	small	medium	negligible	small

Table 3: Descriptive statistics and effect size of the difference between the traditional and the game groups in the immediate knowledge test.

		G1			G2		
		Boys	Girls	All	Boys	Girls	All
Traditional group	Mean	0.67	0.57	0.63	0.61	0.52	0.57
	SD	0.53	0.36	0.47	0.29	0.27	0.28
Game group	Mean	0.59	0.52	0.56	0.83	0.73	0.78
	SD	0.41	0.36	0.39	0.37	0.32	0.35
Effect size	Cohen's d	-0.16	-0.16	-0.17	0.68	0.71	0.67
	effect	negligible	negligible	negligible	medium	medium	medium

Table 4: Descriptive statistics and effect size of the difference between the traditional and the game groups in the retention knowledge test.

		G1			G2		
		Boys	Girls	All	Boys	Girls	All
Traditional group	Mean	-0.61	-0.59	-0.6	-0.01	-0.18	-0.08
	SD	0.51	0.47	0.49	0.41	0.44	0.43
effect size	Cohen's d	-1.17	-1.3	-1.21	-0.03	-0.48	-0.2
	effect	large	large	large	negligible	small	small
Game group	Mean	-0.56	-0.54	-0.55	-0.03	0.03	-0.01
	SD	0.57	0.61	0.58	0.36	0.35	0.35
effect size	Cohen's d	-1.23	-1.15	-1.19	-0.09	0.08	-0.01
	effect	large	large	large	negligible	negligible	negligible

Table 5: Descriptive statistics and effect size of the difference between scores in the retention test and immediate test for the traditional and the game group.

		Overall appeal			Educational value		
		Boys	Girls	All	Boys	Girls	All
Traditional group	Median	3	3	3	3	3	3
	Mean	3.17	3.16	3.17	3.11	2.95	3.04
	SD	0.66	0.69	0.66	0.42	0.52	0.46
Game group	Median	3	3	3	3	3	3
	Mean	3.14	3.23	3.18	2.93	2.7	2.83
	SD	0.44	0.69	0.56	0.27	0.57	0.43
Effect size	Cliff's delta	-0.07	0.041	-0.023	-0.171	-0.226	-0.195
	effect	negligible	negligible	negligible	small	small	small
Wilcoxon's test	statistic	450	200.5	1252.5	442.5	233	1319.5
	p-value	0.5634	0.8113	0.8128	0.0626	0.074	0.0088[a]

Table 6: Comparison of the two groups with regard to the engagement in the lecture. [a]This result approaches significance if Bonferroni correction for multiple comparisons (6 tests in this case) is used.

	Positive judgment to the overall appeal (score 3 or 4)			Positive judgment to the educational value (score 3 or 4)		
	Boys	Girls	All	Boys	Girls	All
Traditional group	93.10%	94.74%	93.75%	96.43%	94.74%	95.74%
Game group	96.55%	86.36%	92.16%	92.59%	75.00%	85.11%

Table 7: Percentages of students who expressed positive attitude towards the seminar and positive judgment of the educational value.

Positive judgment to the game in the game group (score 3 or 4)		
Boys	Girls	All
93.10%	86.36%	90.20%

Table 8: Percentages of students who expressed positive attitude towards the game in the game group.

Appendix A: The learning objective

The micro-game Orbis Pictus Bestialis should be used as a supplement for reinforcing and integrating (in terms of Thomas & Hooper, 1991) knowledge learnt in the expository formal instruction. Alternatively, it can be used as a material for self-learning supplemented by a text material (we have developed) on the topic of animal learning. The game should *not* be used as a stand-alone activity. Thus, the traditional learning format *plus* the game should be viewed as a unit and it is this unit for which it makes sense to specify the learning objectives. In a nutshell, these objectives are:

1) Students are able to explain what is ethology and behaviourism and how these disciplines relate to animal learning; they know about contributions of I. P. Pavlov and K. Lorenz. They are able to explain distinctions between animal research in laboratory and field animal research.

2) Students are able to explain basic terms related to behavioural studies: fixed action patterns, appetitive and consumatory behaviour, Skinner box, etc.

3) Students are able to explain basic terms related to animal learning: imprinting, conditioning, extinction, habituation, and sensitization; positive vs. aversive learning.

4) Students are able to explain major phenomena of positive reinforcement learning, namely paradigms of classical and operant conditioning, and training by means of clicker training, behaviour capturing, shaping, and chaining. They know about the role of reward. They have prerequisite knowledge for training real animals and they should be able, with appropriate guidance, to train real animals using these techniques more efficiently than students receiving

only verbal instruction on this topic. They are able to design a sequence of individual steps to teach an animal a task using positive reinforcing.

It is only the knowledge related to Point (4) that is expected to be reinforced and integrated by playing the game.