Vít Šisler, Michaela Buchtová, Cyril Brom, Zdeněk Hlávka

Towards Empirical–Theoretical Framework for Investigating Learning Effects of Serious Games: A Pilot Study of *Europe 2045*

Abstract

It has been argued that educational computer games are beneficial as educational aid in schools. However, empirical data demonstrating to which extend, if any, computer games outperform so-called "traditional" teaching methods in the context of curricular education are scarce. Moreno's (2005) cognitive-affective theory of learning with media (CATLM) suggests that motivational elements of games can improve meaningful learning by increasing learners' cognitive processing of the to-be learnt information. Yet, at the same time, processing additional extraneous information presented by the game can result in a decline of learning results. To contribute to the development of a general framework for designing and using educational games, our long-term goal is to investigate learning effects of educational computer games in schools in the light of this trade-off. To this end, we aim at carrying out a set of quasi-experimental studies conducted at the school context and using a number of different games as instruments. This paper presents results of our pilot study, in which we tested at five high-school classes (n = 153) whether one particular educational game, *Europe 2045*, can be used as such instrument and whether our method of assessing students' knowledge by classical tests with multiple-choice and open-ended questions is sensitive enough for evaluating the game's learning outcome. The results suggest that (a) *Europe 2045* is an appropriate research instrument; yet, (b) the experimental design should be refined.

Keywords

digital game-based learning, serious games, learning effects, Europe 2045

1 Introduction

Nowadays, educational computer games and serious games are used in multiple fields, including military training, medical and public health training, rehabilitation, and foreign language practising (reviewed, e.g., by de Freitas 2006). Many have argued that educational computer games could also support classical curricular schooling. For instance, 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 and McFarlane 2004). Others propose that games as educational tools may help in 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 difficult to comprehend through factual knowledge only (e.g., Facer *et al.* 2007; Egenfeldt-Nielsen 2005). Thus, educational computer games could help the learners acquire mental models of complicated processes. Additionally, it has been argued that 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).

In our research, we perceive computer games from a broader perspective through the theoretical lenses of Bogost's *procedurality*. In this sense, Bogost (2007, p.vii) has argued that computer games "represent how real and imagined systems work [and] they invite players to interact with those systems and form judgments about them." More importantly, this form of representation is "tied to the core affordances of the computer: computers run processes, they execute calculations and rule-based symbolic manipulations" (p. ix). Similarly, in our research we focus on the procedural aspect of educational games and examine to what extend this "invitation to interaction", i.e. the possibility to navigate oneself in the game's world, manipulate the game's objects, and interact with the game's rule-system, helps to the processes of learning in general and deeper understanding of complicated and interdependent issues in particular. In the light of the main theme of this volume, i.e. applied playfulness, our research aims to expand our understanding of the role computer games, and game as a general principle, could possibly play in our system of formal schooling.

1.1 Evidence about learning effects of serious games

While electronic games gradually enter schools (e.g. Wastiau *et al.* 2009), claims about their usefulness still have not been supported by sufficient empirical findings in the context of curricular education. General reviews of studies comparing instructional effectiveness of games to more conventional forms of instruction reported mixed results (e.g., Randel[1] *et al.* 1992; Hays 2005). The main conclusions of these 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 outcomes; (b) in studies investigating cognitive performance[2] 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 a 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). In addition, most of the games reviewed focus on promoting mathematical or language skills, or physics principles, or economics. Many so-called "educational digital games" are in fact "quizz-based" and/or "drill-and-practice" software, which separates the learning and the gaming part with the game often played as a reward only (which is sometimes bluntly termed "sugar coating learning"). Peculiarly, while the target audience of many games consists of primary school students, the secondary education context (9^{th} till 12^{th} grade in the US system) where many alleged advantages of full-fledged serious games could materialise, is largely ignored (see Brom *et al.* (2011) for details).

Recently, researchers have started to formally investigate how full-fledged serious games can be integrated within secondary schooling system. One possible integration strategy is to design a several hours or days long seminar with various activities organised around playing the game, bypassing the problem of fixed lesson duration. At least two high-school oriented empirical studies (Buch and 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 students claimed that they learned more or at least as much as they usually did. However, these studies did not investigate real learning outcomes.

Another method is to employ PDA devices dedicated only to game-based classes, bypassing problems with accessibility of a computer lab. A prominent proponent of this approach is Klopfer (see Klopfer 2008 for a summary), who generally reports promising outcomes both in affective and cognitive terms, but whose studies tend to employ uncontrolled design and descriptive reports only. 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., Wrzesien *et al.* 2010; Huizenga *et al.* 2009). The latter two studies were comparative and used both qualitative and qualitative measures, and they reported a learning gain, but in Wrzesien *et al.*

1 Randel et al. (1992) also includes non-computer games.

2 The term "cognitive performance" is used here as an umbrella term. Studies investigated whether playing improve various cognitive abilities, ranging from high-level skills to mastering core knowledge of a particular topic.

(2010) game-based activity did not present statistically significant differences with the traditional type of class, and in Huizenga *et al.* (2009) it did, but the control group received a shorter treatment than the experimental group. Finally, one can employ micro-games, which can be played easily on older school computers within school lessons with fixed duration and whose game-play can be designed easily to match parts of the curriculum. While descriptive studies generally suggest that this approach is promising (e.g. Wilensky and Novak 2010), comparative studies report only a slight increment gain in learning compared to the control group or null results (e.g. Wong e*t al.* 2007; Annetta *et al.* 2009; Ioannidou *et al.* 2010; Brom *et al.* 2011).

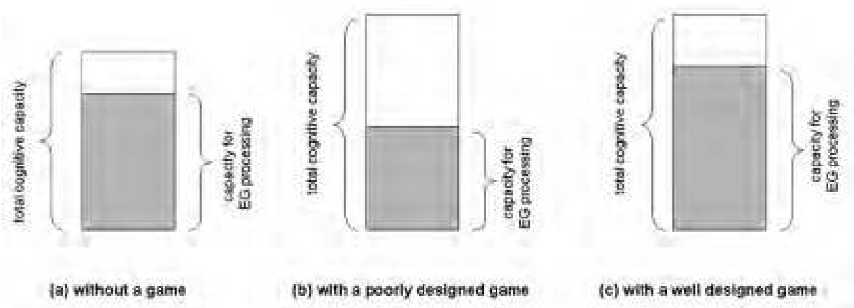
Thus, from the recent studies, we know that serious game based activities are apparently promising, but significant learning gains compared to control groups do not always materialise, which is, generally, in agreement with the older reviews. Additionally, most studies have two drawbacks. First, they tend to investigate gains of a particular game without proposing a theoretical framework as to why the game in question should or should not work, and consequently, how it should be used. As Moreno and Mayer (2007) assert, the development of serious games (and research of their learning effects) should be "grounded in a research-based theory of how people learn" (p.321). Instead of asking whether a computer game promotes learning, it is perhaps more useful to ask which new features it offers and how these features should be employed to promote learning (see (Moreno 2005, pp.2-3; Mayer, 2009, pp.10-14) for more on this issue). Second, most of the studies do not investigate long-lasting effects, but collect data only shortly after the treatment. Arguably, games may hold potential for helping learners to retain whatever knowledge and skills the games teach for longer periods than traditional forms of instruction, as also supported by results of two studies mentioned above that investigated delayed effects of a game-based activity (Wong *et al.* 2007; Brom *et al.* 2011; see also Egenfeldt-Nielsen 2005).

1.2 Serious games in the light of Cognitive-affective theory of learning with media

Computer games are a new technology that brings many new features (some would say "affordances", e.g. Brom *et al.* (2011)) that could promote various types of learning. Our present interest is one particular feature and one type of learning. This "feature" is that complex simulations (as opposed to "drill and practice" edutainment as well as traditional teaching method) enable students to *observe* a computational model of a complex phenomenon, *interact* with it and *actively* inspect its underlying causalities by investigating consequences of their actions. The "type of learning" is mental models acquisition. *Mental models* are typically defined as internal representations of possible behaviour of devices and systems, and possible unfoldments of situations and problems (Johnson-Laird 1983). This includes the depiction of causalities and the ability to draw inferences and making predictions about the reality, including running "internal simulations". This idea is actually not new (cf., e.g., Papert 1993). Note that besides mental models acquisition, a game could teach facts and promote problem-solving skills etc. as discussed above – these are largely out of our present scope.

We further elaborate on this idea in the light of *cognitive-affective theory of learning with media* (CATLM; Moreno 2005; Moreno and Mayer 2007), which is an expansion of Mayer's original *cognitive theory of multimedia learning* (Mayer 2001). In a nutshell, according to the original Mayer's theory, mental models are constructed in the learner's working memory (WM) during *organizing* incoming information and they are later *integrated* with prior knowledge in the learner's long-term memory (LTM) (Mayer 2001; note that WM and LTM are classical psychological concepts, see, e.g. Baddeley *et al.* 2009). This is called *generative cognitive processing*. Additionally, the information must somehow appear in the WM at the first place; that happens when the learner *selects* it actively from the learning material. This is called *essential cognitive processing*. Together, "essential and generative processing result in the creation of a meaningful learning outcome" (Moreno and Mayer 2007, p.315). When higher *cognitive capacity* is available for essential and generative processing (EG processing), better the mental model is integrated with prior knowledge. Sadly, total cognitive capacity of a learner does not equal the cognitive capacity for EG processing. The CATLM actually assumes the following trade-off. On the one hand, motivational factors can increase total cognitive capacity (or lack of a learner's motivation may fail to engage the learner in EG processing even when cognitive capacity is available). On the other hand, capacity available for EG processing may be reduced by processing of *extraneous details* that are not directly related to the learning content but may, for instance, make the game engaging, thereby contributing to increasing total cognitive capacity of the learner or recruitment of EG processing (Fig. 1).

To contribute to the development of a general framework for designing and using complex serious games, our long-term goal is to investigate learning effects of serious games in schools in the light of this trade-off. To this end, we aim at carrying out a set of quasi-experimental studies conducted at the school context and using several different games as instruments. In these studies, we eventually aim at manipulating two factors: (a) the level of engagement by compromising the gaming element, and (b) the extraneous details by altering interfaces of the games. We have already developed two serious games that have been successfully integrated within formal secondary education system, *Europe 2045* (Brom *et al.* 2010) and *Orbis Pictus Bestialis* (Brom *et al.* 2011). We have also demonstrated "medium positive effect size" concerning learning gains for the latter game (comparing to the control group; Cohen's $d = 0.67$ (Cohen, 1988)). Additionally, we have been also developing a new complex serious game as an aid for teaching history of the Czech lands in the 20th century at high schools. Thus, we potentially have research instruments available.



the game (a) than with the game (b, c) due to the game's motivational elements, capacity for useful learning, i.e. for EG processing, can be lower with the game (b).

1.3 Purpose of the present study

The goal of this paper is to present results of our pilot study that aimed at investigating whether *Europe 2045*, embedded within a teaching framework, can be used as one of such research instruments for investigating learning effects of serious games in schools. The educational goal of the game is to teach:

- *facts* about European states, EU institutions and policies and typologies of political inclinations;
- *mental models* concerning large-scale processes and socio-political notions, such as a model of "energy dependence" or "liberalism"; and
- several *high-level skills*, including the ability to discuss and work in teams.

Concerning our theoretical framework, presently, we are interested in mental models only; however, if the game compromises learning of factual knowledge, it cannot be recommend as an educational aid. Thus, to accept the game as a research instrument, it is necessary that:

a) it teaches facts at least as good as traditional forms of instruction,
b) it is detectable that it teaches mental models.

Question (a) can be answered using classical comparative study with quantitative measures. Question (b) does not require comparative design. It can be answered using a combination of qualitative approach (e.g., interviewing, examining) and quantitative study with pre-test/post-test design.

There is, however, a problem with using tests as measuring instruments. While the knowing of facts can be assessed by classical tests easily, as repeatedly demonstrated in curricular schooling, it is less clear whether this approach is appropriate, most notably sensitive enough, for assessing quality of acquired mental models. Note that our situation is complicated by the fact that we conduct field research directly in schools and not in a laboratory, which means that a higher noise in the data can be expected.

Thus, this study has also another goal, to answer the question:
c) are our mental model tests sensitive enough?

In particular, we aim at assessing whether two types of questions, multiple choice and open-ended questions, yield detectable differences between the game group and the control group. Negative and positive differences mean that the tests are sensitive and the game is detrimental or beneficial to mental models acquisition, respectively. Null results mean either the game is comparable to traditional forms of instruction or the tests are unable to detect differences. In all cases, the game can be used as an instrument provided we are able to manipulate extraneous details of the game (which we can; see Fig. 1), but in the case of null results, we should also seek another measuring method. In addition, concerning Points (a) and (c), we are interested both in immediate and delayed effects.

In this paper, we present results of our pilot study elucidating these questions. Section 2 introduces *Europe 2045*. Section 3 details the experimental design. Section 4 presents the results and Section 5 discusses the outcomes and further research.

2 Europe 2045

Europe 2045 is played in teams. Each student represents a member state of the European Union and the whole class represents the EU. The game can be played between 8 and 24 students, while the teacher assumes the role of coach/tutor. At the beginning of the game session, the game situation closely resembles the real state of affairs in Europe in the end of the first decade of 21st century. The game proceeds in rounds with each round representing one year. The game also features an in-game encyclopaedia (Fig. 3). This structured, hypertext-linked set of web-pages provides supplementary information, which is both relevant for success in the game and which summarises related real world information.

Europe 2045 combines the principles of two game genres: a multi-player online computer game and a social role-playing game. Note that the latter is not played on computers only but in the classroom and the student's "real" space as well. Both games are interconnected, which is a key feature of *Europe 2045* (see Brom *et al.* 2010, for details).

Europe 2045 features three layers of game-play: the economic layer, the diplomatic layer, and the storytelling layer. On the *economic layer*, every student defines the domestic policy of his/her state beginning with tax levels and environmental protection and graduating on to issues such as the legalisation of same-sex marriage and privacy protection (Fig. 2).

On the *diplomatic layer*, the player has an opportunity to present drafts for policy changes to the EU (for issues such as common immigration policy, stem-cell research or agricultural quotas). The discussions about these changes take place in the classroom (Fig. 3), where the teacher, who also gives short lectures to contextualise gaming issues, moderates them. The discussions can be conceived as simulating negotiations at a wide array of EU institutions, such as The European Parliament, The European Council and The Foreign Affairs Council.

Every player has his/her own project to try to push through at the European level. A project is essentially a vision of how the EU should look in the future (e.g., the Green Europe project supports environmental protection and investment into alternative power resources, while the Conservative Europe project strives to preserve traditional values). At the same time, a project is formally defined by: (a) a set of policies that should be put in force, (b) a set that should be suspended, and (c) a set to which the project is indifferent. From the gaming perspective, these projects present roles the students can play. Because some projects agree or disagree upon the same subset of policies, each player can find a teammate to support his/her intended particular policy change. The final appearance of Europe at the end of each game session is thus the result of intense negotiations and voting in a given player group.

On the *storytelling* layer, players face various simulated scenarios and crises relating to key contemporary issues that unified Europe faces (such as the humanitarian crisis in Darfur). The players must react to all these events and, in cooperation with fellow players, seek appropriate solutions. During the course of the game, the students typically witness the short- and/or long-term effects of their decisions.

The game was designed to support two modes of play: during regular classes over the term, or during a special one-day seminar. In this study, we employed the former mode. As detailed later, students played the game twice a week over four weeks of regular education; every session lasted 45 minutes (one school unit) and amounted to one game round. Because *Europe 2045* is an asynchronous client-server game, the students could also negotiate at other times via the in-game on-line forums. Usually, they prepared for negotiations and played the game outside classes for approximately an hour a week.

The game was released in 2008 and it has been was played by more than 2500 students since then. The evaluation study (Brom *et al.* 2010) suggested that it was successfully integrated within formal schooling systems (e.g., out of 188 students from that study, 37% rated the game as excellent and 41% as good; only 6% of all the answers were negative).

Fig. 2. Europe 2045 screenshot: GUI of inner politics settings.

Fig. 3. The in-class discussion within Europe 2045 (Courtesy of Ivo Šebek, used with permission.)

3 Study design

The study was designed to answer Questions (a) – (c) from Sec. 1.3, focusing on mental models acquisition. The study was conducted in spring 2011 at two Czech high schools. One school was used for calibrating tests (2 classes, n = 59) and one for the real study (5 classes, n = 153; see also Tab. 1).

3.1 Experimental design

The study compares a teaching module on the topic of European Union and political systems that used *Europe 2045* (game classes) against a comparable module using a more traditional teaching style (control classes). The game classes received eight lessons (1 introductory, 6 with the game, 1 final lesson). The control classes received eight lessons. Every school lesson lasted 45 minutes.

Pre-tests were administered a week before the introductory lecture in both groups (in March 2011), including questionnaires eliciting biographical and ICT-literacy data (for both types of classes). Post-tests were administered the next lesson after the eighth lesson (April or May, depending on the school schedule), the experimental group also received questionnaires evaluating the simulation *Europe 2045*. Within three weeks after the immediate post-tests, three focus groups were conducted with students from experimental groups. Delayed post-tests were administered in September 2011, i.e., after the summer holidays. Every test session lasted about 45 minutes, students from one class were given the test at the same time and they undertook the test in the same room, i.e. their regular classroom. Students received credits for the immediate post-test, but not for the other tests. The students were informed in advance neither about the pre-test nor the delayed post-test, but they knew about the immediate post-test.

3.2 Participants, sampling

We worked with five classes at the experimental school and two classes at the calibration school. The students were from 15 to 17 years old (11^{th} grade in the US system). Numbers of students in each class at the experimental school are given in Tab. 1.

It was impossible to divide every class into a control group and a game group due to school constraints. Instead, at the calibration school one class was chosen as a control class and one as a game class randomly. At the experimental school, two classes specialised in humanities, two in natural sciences and one in computer science. One humanities class and one natural sciences class were chosen as control groups randomly. The other three classes played the game.

For the purposes of the qualitative study, we formed three focus groups, each involving eight students – four girls and four boys – all of them from experimental classes. The participants were chosen on their level of activity in the simulation, which was operationalized as the number of voting of a particular student in the game: four students voting in more than three quarters of the rounds, two in approximately half of the rounds, and two in less than one quarter of the rounds.

	Total (% of boys)	Pretest	Immediate	Delayed
H.e	32 (56%)	24 (58%)	31 (55%)	27 (59%)
H.c	30 (34%)	27 (39%)	27 (37%)	18 (33%)
N.e	32 (61%)	30 (63%)	31 (61%)	26 (65%)
N.c	31 (55%)	31 (55%)	28 (54%)	25 (52%)
IT.e	28 (82%)	23 (78%)	24 (83%)	25 (84%)

Tab. 1. Numbers of students in classes, and numbers of students participating in tests. "H" stands for humanities, "N" for natural sciences, "IT" for computer sciences, "c" for control, "e" for experimental.

3.3 Teaching methodology

Every lesson with *Europe 2045* had two parts; the discussion (about 25 minutes) and the teacher's lecture (about 20 minutes). The discussion focused on in-game events and EU policies (see Sec. 2). The events were selected in advance according to the school curricula. The political propositions included immigration politics, cultural quotas and environmental taxes. In his 20 minutes, the teacher summarised the discussion, clarified problematic points and contextualised them. In every lesson, he also briefly explained one topic connected to the EU, such as membership criteria of the EU. It is important to note, that we do not separate the game related learning results from learning through discussing the game and in-game events. Discussion is an integral and inseparable part of the gameplay of *Europe 2045*. At the same time, debriefing is known to be an important phase in using simulation games (e.g., Peter and Vissers 2004) and some suggest that it should be used in computer games and instructional programs as well (e.g., Hays 2005).

The syllabus followed by the control classes roughly corresponded to the syllabus of the experimental classes, and the teaching method followed a similar pattern as well. The discussions were replaced by short students' presentations. Topics for these presentations were selected randomly from a pool of topics that were connected to the in-game events and the policies the students discussed in the experimental classes. The teacher's lectures also corresponded to lectures in the experimental classes as closely as possible.

At the beginning of the learning period, students from both groups received the same didactic text, the printed in-game encyclopaedia of *Europe 2045* covering the main themes about the European Union.

3.4 Questionnaires and tests

The knowledge tests were actually investigated artefacts (see Sec. 1.3); we had three of them: pre-, post- and delayed knowledge tests. Additionally, pre-tests were given with questionnaires eliciting biographical and ICT-literacy data (for both types of classes) and immediate post-tests with questionnaires eliciting data on subjective evaluation of the experience with the game (for the game classes only). All questionnaires were constructed by the authors (V.Š., M.B., C.B.).

Data from the background questionnaires will not be analysed here; suffice is to say that they showed that all students except two used a computer for more than one hour a week and the majority for more than five hours a week.

Subjective assessment questionnaire contained 16 questions; for present purposes, only two of them are analysed:

1 *Mark which answer most closely corresponds to your opinion: (a) I liked the game and I would like to continue in playing; (b) I liked the game but I do not want to continue in playing; (c) I did not like the game and I do not want to continue in playing.*
2 *Rate the overall appeal of the following three parts of the game: (a) discussions and negotiations; (b) economic model; (c) encyclopaedia readings* (Likert items with 4-point Likert scale).

The most important are the knowledge tests. The test questions were designed to match the curriculum of the teaching module and discussed with two social science teachers. Each knowledge test consisted of three groups of questions. The order of the questions was the same for all subjects and the questions from different groups were mixed together. Questions were paired across the tests; paired questions tested the same knowledge.

The first group comprised five factual multiple-choice questions. The second and the third part consisted of multiple-choice (6) and open-ended (4) questions, respectively, that assessed students' understanding, i.e. "the ability to construct a coherent mental representation [model] from the represented material" (Mayer 2001, p.19). Essentially, these so-called "transfer questions" tested the students' ability to use the information from the lessons in novel situations. Design of these questions was inspired by Mayer's approach (2001, p. 39) and the questions eventually roughly corresponded to Mayer's conceptual, prediction and redesign types of questions (p. 39) (Fig. 4). The questionnaires were calibrated on the "calibration" school.

After completion, two independent experts scored the open-ended questions on 11-levels ordinal scale. Spearman correlation coefficient between the two scoring persons was high, > 0.9. The tests were not anonymous.

A) The current immigration policy of the European Union comprises primarily the following:
a) policies against illegal immigration into the EU
b) harmonization of regulations enabling employment of qualified foreigners from non-EU countries
c) active integration of immigrants into host societies in the EU
d) relocation of foreigners from the EU

B) Write down 5 words or groups of words, which according to your own opinion most accurately describe the weaknesses and shortcomings of the current decision-making procedures of the EU: ....
....

C) EU as a whole is challenged by low natality rates and a decreasing share of citizens in the productive age. Which of the below-listed solutions has according to your own opinion the highest chances to help to improve the above-mentioned conditions in the near future?
a) unification of social benefits in the member states of the EU
b) increase of the retirement age in the member states of the EU
c) active support of immigration and acceptance of migrant workers from non-EU countries
d) decrease of tax burden for citizens in the productive age

Fig. 4. Question examples: (a) factual, (b) openended, conceptual transfer question, (c) multiple choice, redesign transfer question.

3.5 Focus groups

Generally, qualitative approaches can provide a more subtle understanding of learning processes and motives than quantitative methods. Focus groups are a particular qualitative method, which is used throughout the social sciences (e.g., Morgan 1996) as a "specific research technique of series of discussions designed to study the respondents' needs and feelings within specific areas" (Krueger 2001).

The main objectives of our focus groups were to: (1) ascertain students' opinion about differences between "drill and practice" methods and educational simulations, focusing on mental models acquisition; (2) identify the elements helping in the process of learning or strengthening the motivation towards learning; (3) provide the understanding of the students' acceptance of educational simulations at school.

Accordingly, the group discussions were divided into three sections, (A) the attitude towards the educational simulations and games, (B) the attitude towards the simulation used during the experiment, (C) the comparison of the simulation with a classic lecture. Part (C) is directly related to Objective (1), our present interest.

During the group discussions, the students were also given a questionnaire related to Objectives (1) – (3) with six Likert items with 10-points Likert-scale. Each focus group took approximately 45 minutes.

3.6 Data analysis

For every evaluated knowledge test, we calculated three average scores corresponding to its three parts: factual, multiple choice transfer, and open-ended transfer questions. We thus obtained three scores in the pre-test, the immediate post-test and the delayed post-test for every student, unless a student missed a test. All students participating in at least one test were included in the analysis.

The effect of *Europe 2045* on each of the three types of students' average scores was assessed by a mixed linear model (Pinheiro and Bates 2000) with random effect of class and random effect of student nested within the class accounting for the within-class and within-student correlations (i.e. correlations between each student's repeated test results). The fixed effects included the effect of gender, the effects of the traditional teaching method on the average score observed in the immediate and the delayed post-test (compared to the result in pre-test), and the effect modifications caused by playing *Europe 2045* in the immediate and the delayed post-test. We assumed that any effect of playing *Europe 2045* in the pre-test did not exist.

The statistical analysis was performed in the statistical computing environment R (R Development Core Team, 2011) with library *nlme* (Pinheiro *et al.* 2011).

Within the qualitative data analysis, we identified a thematic framework of each statement, indexed and mapped the frequency and development of the thematic framework and finally interpreted the findings in the framework of the Objectives (1) – (3). The participants' and data triangulation verification was ensured by (a) collaboration of two focus group moderators who independently preceded the group interviews, and (b) by a subsequent collective debriefing, including the research leaders, where the findings were discussed before the analysis itself.

4. Results

4.1 Do students like the game?

The overall subjective evaluation of the game was positive. From 84 students participating in the subjective evaluation questionnaire, 74 (87%) students liked the game, from which 21 (28%) would like to continue playing the game. Only 10 (11%) students disliked the game. The most interesting part of the game was interfacing with the simulation (Mean = 1.98; SD = 0.87; 1 = interesting; 4 = uninteresting), followed by discussions and negotiations (Mean = 2.32; SD = 0.89) and encyclopeadia reading (Mean = 2.83; SD = 0.93). This is in general agreement with our previous results (Brom *et al.* 2010).

4.2 Focus groups

The outcomes of the qualitative study were extensive. We present here only the part connected to the theme of mental models acquisition.

During the focus groups, the respondents commented on the differences of the knowledge gained through an educational simulation and a classic lecture. Approximately 1/3 of the students spontaneously declared this fact and in the most cases the rest of the group agreed on it afterwards. They described this difference mainly by the words such as "*understanding*", "*better representation*" and "*view from inside*". Therefore, the classic elements of simulations – clear representation of complex processes, a direct feedback, and visualization or modelling of the inner relations of a system and/or processes – apparently help in developing mental models.

Quotations:

"The simulation taught me to think politically, economically. In the lesson you learn definitions, that's the difference. In the simulation you wonder how it works."

"Maybe I more understand the connections. If I do something in the country, what could be the consequences."

"I was representing Romania. I cut there the taxes and the social support. It brought various strikes and more homeless people. Now if I hear something like that in the news, I can better imagine the situation."

"It was not a classic content. Normally we don't learn this at all."

In contrast, in the questionnaires distributed during the focus groups, one question asked the students to compare the amount of knowledge gained through a classic lecture and an educational simulation. Within the simulation, the average value was around 5.8 (SD = 1.9; on the scale from 1 = the least, 10 = the most) and within the classic lecture around 7.3 (SD = 1.5).

From the following discussion it emerged that the students perceive simulations mainly as a tool for practicing and strengthening the knowledge gained by classic lecture or studying text or on-line sources. This outcome can be influenced by the strong schooling tradition in the Czech Republic where the curriculum mainly focused on factual knowledge and "drill and practice" methods.[3]

Quotations:

"The classic lecture is good to learn facts, in the simulation we can practice it."

"The simulation was good because it brought us to the situation which we cannot experience in the real life. It is more about the feeling to try it and also to enjoy. When you try it you can remember it."

"The game was good to practice and to verify if we really understand it."

Analysis of frequent statements suggested that respondents recognized the educational simulation as a tool dissimilar to classical lectures mainly because the simulation offers an interactive experience and provides feedback. This may be the reason why students expected different learning effects, in particular, practising and deeper understanding

4.3 Learning effects

The results are summarized on Fig. 5. The statistical analysis was performed by fitting a linear mixed effects model for each average score (i.e., concerning the factual, multiple-choice transfer, and open-ended transfer questions). The effect of gender was not significant (p-values=0.69, 0.50, and 0.28 for factual, mental transfer and open-ended questions, respectively). Concerning the factual score across all classes, we have observed a significant increase of the average score in the immediate post-test (0.32, p-value<0.0001) with a correction towards the pre-test result in the delayed post-test (0.13, p-value=0.002). The estimated effects of *Europe 2045* on the average factual scores were negligible (-0.03 and 0.02 in the immediate and the delayed post-test) and not statistically significant (p-values 0.47 and 0.69).

For the multiple-choice transfer average score, we have observed a negligible improvement in the immediate post-test (0.04, p-value=0.079) and a somewhat larger statistically significant improvement in the delayed post-test (0.14, p-value<0.0001). The effects of *Europe 2045* were again negligible and not statistically significant (estimates 0.04 and 0.02, p-values 0.15 and 0.47 in the immediate and delayed post-test, respectively).

For the open-ended questions, we have observed an improvement in the im-

3 There has been being an on-going school reform in the Czech Republic for the last about ten years.

mediate post-test (0.13, p-value<0.0001) and a somewhat smaller improvement in the delayed post-test (0.05, p-value=0.0175). The effect of *Europe 2045* was significantly negative in the immediate post-test (-0.08, p-value=0.0004) and non-significant in the delayed post-test (-0.02, p-value=0.41). A closer look at the data reveals that the negative effect is caused mainly by very poor results of one particular class, i.e. H.e. in the immediate post-test.

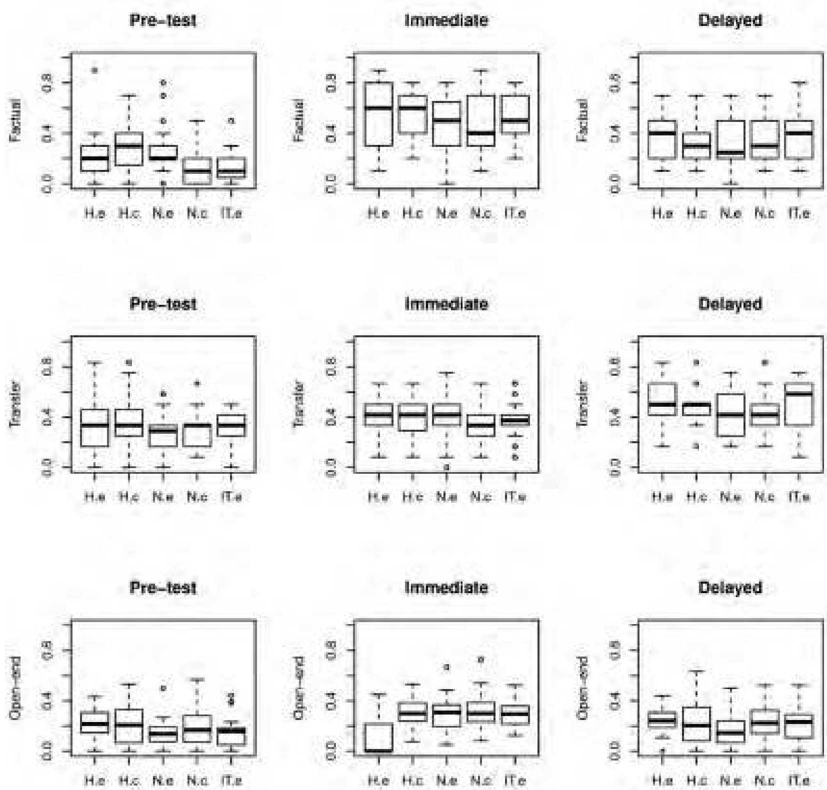


Fig. 5. Box-plots of average students' scores for factual, multiple choice transfer, and open-ended transfer questions in the pre-test, in the immediate post-test, and in the delayed post-test. The box-plots provide a simple graphical view of the median, quartiles, maximum, and minimum of test results within each class. The ends of "whiskers" denote the lowest/highest datum still within 1.5 of the respective interquartil range. The dots represent outliers.

5. Discussion and conclusion

The results showed that:

(1) Students, in general, appreciate the game. They particularly like the simulation and, albeit to a lesser extent, the discussions. These are the parts of the game that can potentially promote mental models acquisition and students seem to interface with them.

(2) Across all tests, there are significant, or nearly significant differences between pre-test scores and immediate post-test scores, and pre-test scores and delayed post-test scores. This suggests that both types of treatments have an effect on students' knowledge and we can somehow measure this effect. However, the difference is the largest for factual knowledge questions and smaller for questions assessing mental models acquisition. The fact that the game teaches factual knowledge is not very surprising because (a) this is one of the game's educational objectives, and (b) the game is supplemented by a teacher's inputs. A low between-tests difference for mental models questions is troublesome: this either means neither treatment is very effective in this regard, or that our tests are not very sensitive, or that the tests measure something different that the treatments teach.

(3) The game supplemented by the teacher's contextualization and short lectures teaches factual knowledge at least as good as comparable, traditional forms of instruction, as measured by our factual test questions.

(4) The game is most likely as good as the traditional teaching module concerning mental models acquisition, as measured by multiple choice and open-ended questions. The significant negative effect of *Europe 2045* in immediate post-test's open-ended questions is clearly caused by the disproportionally poor performance of H.e class in this group of questions in this particular test. The reasons for such performance of H.e are more or less unknown to us, because this class otherwise performed well and the immediate test was marked, which makes "low stake test" interpretation improbable (note we double-checked the data). The reason seems to be connected to the way the tests were administered within the surrounding context, and the design of the tests. The poor performance is unlikely the effect of the game due to the disproportionally good performance of H.e in delayed tests.

(5) At the same time, the qualitative data suggest that students positively think they learned a kind of knowledge that can be conceived as mental models (even though the Czech students tend not to call this type of learning "knowledge acquisition"). This again adds to the suggestion that the game has certain "learning effect" but we do not measure it appropriately.

Regarding our research questions (Sec. 1.3), the results show that the game is suitable as a research instrument for our intended set of larger studies (due to

Points (1), (2), (3), (5)) but our tests may either not be sensitive enough along the mental models acquisition axis or measure something the treatments do not really teach (Points (2), (4), (5)). Thus, we are presently considering a different experimental design, where the game will be played within one day workshop out of schools (to address the "H.e issue" and also for generally better controlling the intervening variables), and, most importantly, a different way of measuring mental models acquisition, including oral examining the students and engaging them in a problem solving activity where they should practically demonstrate their knowledge. As soon as we have new measuring instruments in hand, the plan, following our long-term vision, is to manipulate the level of engagement of the game to investigate whether this has an effect on learning.

In a more general perspective, one may ponder on whether teaching with *Europe 2045* is better than traditional teaching approaches. Similar to most other empirical game-based studies, our data are inconclusive in this regard. However, we would argue that this is an ill-posed question. The point is to investigate what features of games can contribute to learning, not whether a particular game, played in a particular context and in a particular way, is better, or not, than a different type of instruction. Generalizing answers on these questions is highly problematic. Instead, our purpose is different: we aim at investigating whether "the magical something" that makes games so engaging, i.e. their ludic element, contributes to learning and under what conditions it does so. In this regard, the present study merely, yet importantly, assessed usability of our experimental design and *Europe 2045* as a research instrument for investigating the learning effects of serious games.

Acknowledgement

The paper was supported by the project NAKI "Příběhy z dějin československého státu: výzkum a experimentální vývoj softwarových simulací pro výuku historie českých zemí ve 20. století", No.: DF11P01OVV030, financed by the Ministry of Culture, Czech Republic, and investigated at the Faculty of Arts and the Faculty of Mathematics and Physics of Charles University in Prague and the Institute of Contemporary History of the Academy of Sciences of the Czech Republic in 2011–2014.

Literature

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.

Baddeley, A.D., Eysenck, M., and Anderson, M.C. (2009) *Memory*. Hove: Psychology Press.

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.

Bogost, I. (2007) *Persuasive Games: The Expressive Power of Videogames*, Cambridge: MIT Press.

Brom, C., Preuss, M. and Klement, D. (2011) Are Educational Computer Micro-Games Engaging And Effective For Knowledge Acquisition At High-Schools? A Quasi-Experimental Study. *Computers & Education* 57, pp.1971-1988.

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 effect of 'Global Conflicts: Middle East'. In: Santorineos, M. and Dimitriadi, N. (eds.) *Gaming Realities: A Challenge for Digital Culture*. Fournos, Athens 93 – 97.

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

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

Facer, K., Ulicsak, M., Sandford, R. (2007) Can Computer Games Go to School? *Becta. Emerging Technologies for Learning. British Educational Communications and Technology Agency*, Coventry [Online]. Available at: http://partners.becta.org.uk/index.php?section=rh&catcode=_re_rp_ap_03&rid=11380 [Accessed: 6th September 2009].

de Freitas, S. (2006) Learning in immersive worlds. *Joint information systéme committee*. [Online]. Available at: http://www.jisc.ac.uk/eli_outcomes.html [Accessed 16th November 2010].

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

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.

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.

Johnson-Laird, P.N. (1983) *Mental models*. Harvard University Press.

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

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

Krueger, R.A. (2001) Designing and Conducting Focus Group Interviews. In. *Social Analysis: Selected Tools and Technics.* Washington: Social Development Department, The World Bank.

Krueger, R.A. and Casey, M.A. (2000) *Focus Groups: A Practical Guide for Applied Research.* Sage Publications.

Mayer, R.E. (2001) Multimedia learning. New York: Cambridge University Press.

Morgan, D. (1996) *Focus Groups as Qualitative Research.* Sage Publications, Inc.

Mayer, R. E. (2009). *Multimedia learning* (2nd ed). New York: Cambridge University Press.

Moreno, R. (2005). Instructional technology: Promise and pitfalls. In L. Pytlik Zillig, M. Bodvarsson, and R. Bruning (Eds.) *Technology-based education: Bringing researchers and practitioners together* (pp. 1–19). Greenwich, CT: Information Age Publishing.

Moreno, R. and Mayer, R. (2007) Special Issue on Interactive Learning Environments: Contemporary Issues and Trends. *Interactive Multimodal Learning Environments*. Springer Science + Business Media.

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

Pinheiro, J.C. and Bates, D.M (2000). *Mixed effects models in S and S-Plus*. Springer, New York.

Pinheiro, J.C., Bates, D.M. DebRoy, S., Sarkar, D. and R Development Core Team (2011) *Linear and Nonlinear Mixed Effects Models*. R-package version 3.1-102.

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.

R Development Core Team (2011) R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria.

Sandford, R., Ulicsak, M., Facer, K. and Rudd, T. (2007) *Teaching with Games. Using commercial off-the-shelf computer games in formal education.* Futurelab. Bristol. [Online]. Available at: http://www2.futurelab.org.uk/resources/documents/project_reports/teaching_with_games/TWG_report.pdf [Accessed: 10th December 2011]

Squire, K. and Steinkuehler, C. (2005) *Meet the Gamers. Library Journal.* [Online]. Available at: http://www.libraryjournal.com/article/CA516033.html [Accessed: 10th December 2011]

Wastiau, P. *et al.* (2009) *How are digital games used in schools: Complete results of the study*. European Schoolnet. [Online]. Available at: http://games.eun.org/upload/gis-synthesis_report_en.pdf [Accessed: 10 December 2011].

Wilensky, U. and Novak, M. (2010) Understanding evolution as an emergent process: learning with agent-based models of evolutionary dynamics. In Taylor, R. S., 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: pp. 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.