You Like It, You Learn It: Affectivity and Learning in Competitive Social Roleplay Gaming

Cyril Brom, Vít Šisler, Michaela Slussareff, Tereza Selmbacherová, Zdeněk Hlávka

Faculty of Mathematics and Physics, Charles University in Prague

This work has been accepted for publication in the **International Journal of Computer-Supported Collaborative Learning** (Received: 29 December 2014 /Accepted: 27 June 2016 / First Online: 29 July 2016). The present text is a preprint. The final publication is available at Springer via http://dx.doi.org/10.1007/s11412-016-9237-3.

Source of funding

This research was funded by Czech Grant Science Foundation (GA ČR) (Project nr. P407/12/P152).

Acknowledgement

We thank all research assistants and lecturers who helped to conduct the experiment, namely: T. Holan, I. Šebek, J. Vlasák, L. Kolek, V. Dobrovolný, M. Grecká, M. Denemarková, J. Tupá, O. Smíšek, R. Římanová, J. Lacka, M. Abrahámová, J. Fiala, M. Hampacherová, K. Pavelka, T. Stárková, V. Šálený, J. Volák, V. Zemanová, T. Pospíšil, D. Wagner, S. Feyglová, K. Vávrová, J. Janovský, V. Šnoblová, E. Bednaříková, O. Šíp, I. Pecháčková. We thank Jan L. Plass, Jan Lukavský, and Sidney D'Mello for discussing this work with us. We thank all high school teachers enabling us to conduct the experiment. We also thank three anonymous reviewers and the editor Rolf Steier.

The human data were collected with APA ethical principles in mind.

Abstract

Despite the alleged ability of digital game-based learning (DGBL) to foster positive affect and in turn improve learning, the link between affectivity and learning has not been sufficiently investigated in this field. Regarding learning from team-based games with competitive elements, even less is known about the relationship between competitiveness (as a dispositional trait) and induced positive affect. In this media comparison study with between-subject design, participants (N = 325; high school and college students) learned about the EU's policy agenda by means of a debate-based method delivered through one of three educational media: a) through a social role-playing game with competitive elements played on computers, b) through a very similar game played without computers and c) through a non-game workshop. Unlike many previous DGBL studies, this study used participant randomization and strived to address the teacher effect and the length of exposure effect, while also using the same learning materials and a very similar educational method for all three treatments. Both games induced comparatively higher generalized positive affect and flow. Participants also learned more with the games. Positive affect, but not flow, mediated the influence of educational media on learning gains. Participants' competitiveness was partly related to positive affect and experiencing flow but unrelated to learning gains. These outcomes held both when the game was played using computers, as well as without them. The study indicates that the ability of an educational intervention to instigate positive affect is an important feature that should be considered by educational designers.

Keywords: positive affect, flow, learning effects, educational games, role-playing, competition

Article

1. Introduction

Digital games for learning (also called serious games or educational videogames) are being used more frequently in schools, and their learning advantages are increasingly subject to investigation. Dozens of comparative studies examining the learning effects of games relative to the learning effects of "traditional" types of instruction are now available. The key meta-analyses of these studies (Sitzmann, 2011; Wouters, van Nimwegen, van Oostendorp, & van der Spek, 2013) reported a modest superiority of digital game-based learning (DGBL) in cognitive terms; however, the latter meta-analysis also found that, in studies with randomization, this effect (favoring game-based learning) disappears (p. 259). This means that at least part of games' superiority over more traditional methods can be explained by the poor design (i.e., without randomization) of many studies.

One of the important arguments for backing the DGBL approach is that games, in general, are motivating (e.g., Malone, 1981; Garris, Ahlers, & Driskell, 2002). This raises hopes that when digital games are employed in the service of learning, learning outcomes will flourish. However, if a game is *actually* used for learning, is it really more motivating than "traditional" instruction? Contrary to common expectations, this is not a given. This question has also not been researched sufficiently. Affective and motivational variables, such as generalized positive affect, flow, and intrinsic motivation, were only investigated in about one-third of media comparison studies conducted so far (see Vogel et al., 2006; Sitzmann, 2011; Wouters et al., 2013). Very few studies have directly investigated the relationship between affects/motivations and actual learning outcomes in the DGBL context (we are aware of: Adams, Mayer, MacNamara, Koenig, & Wainess, 2012; van Dijk, 2010; Iten &

Petko, 2014; Ritterfeld, Shen, Wang, Nocera, & Wong, 2009; Gianakos, 2013; Sabourin & Lester, 2014; Stege, van Lankveld, & Spronck, 2012; see also Habgood & Ainsworth, 2011). Some of these studies indeed reported controversial results (e.g., Adams et al., 2012; Iten & Petko, 2014; Stege et al., 2012).

One of the reasons why digital games, in general, may be motivating and/or induce positive affect and flow is that (many of them) feature competitive elements. However, conflicting results were reported regarding the impact of competition in the context of DGBL (see Vandercruysse, Vandewaetere, Cornillie, & Clarebout, 2013; Plass et al., 2013; DeLeeuw and Mayer, 2011; but also Ke, 2008; ter Vrugte et al., 2015). This could be caused simply by learners' different attitudes towards competitive situations. However, to the best of our knowledge, no DGBL study has investigated the link between competitiveness, as a dispositional trait, and positive affect or flow induced by playing a serious game with competitive elements. This is an underexplored issue.

DGBL visionaries made bold claims about the necessity to augment formal schooling systems with digital games (summarized in Mayer, 2014a, pp. 13-15). The reason often mentioned by them is that today's adolescents and young adults, i.e., the "digital generation" (Prensky, 2001), have grown up in digital world and are thus insensitive to "classical" non-digital education (Tapscott, 1998, p. 131; Prensky, 2001, p. 1). Despite this, acceptance of digital technologies by students *at school* is not guaranteed (e.g., Bourgonjon, Valcke, Soetaert, & Schellens, 2010; Courtois et al., 2014). At the same time, the issue of comparing digital games to their non-computer counterparts remains unaddressed. Would learning be enhanced or hindered if students played a game using a digital technology vs. "old-fashioned" pen and paper?

Considering all the points above, this study addresses the following research questions:

a) Assuming that we deliver particular educational content through a particular educational method via a non-game workshop vs. game-based education, i.e., through different *instructional media*, would the instructional medium influence positive affect, flow levels and learning outcomes?
b) Does competitiveness (i.e., a participant trait) moderate the effect of media on positive affect and flow levels?
c) Do positive affect and flow levels induced by the instructional medium mediate the influence of the medium on learning outcomes?
d) When we use two game-based media, one employing computer technology and the other using pen and paper, would the type of "delivery" technology influence positive affect, flow level, and learning outcomes?

This study's primary research question is (c). While it may seem that all of these research questions should have been answered long ago, only small steps have actually been taken to answer them. To address these questions in the current study, learners learn a certain topic from a specific debate-based educational method embedded in one of the following three media: a computer game, a non-computer game, and a non-game workshop (i.e., a between-subject design; see Table 1). The educational method is very similar across all the conditions and the topic is exactly the same (i.e., the EU's policy agenda). The sample consists predominantly of high school students. We purposefully use *Europe 2045* as the research game (Brom, Šisler, & Slavík, 2010), one that we had developed in the past. There are four reasons for this. First, it is a team-based game and there is some initial evidence that games played in dyads or larger groups are particularly effective for learning (Wouters et al., 2013; p. 258). Moreover, this game has been successfully implemented and used in more than one hundred high schools in the Czech Republic; indicating that it is a promising intervention to address our research questions. Second, it is a social role-playing negotiation game with game mechanics similar to those of some other educational games (e.g., Mochocki, 2013). This is important from a practical perspective: this study's outcomes can be straightforwardly generalized to similar games. Third, the game features elements of mild competition, which is needed to address Research Question (b). Fourth, it uses a specific educational method, which can also be delivered by a non-game medium and by an equivalent non-computer educational game (this enables us to create closely comparable learning experiences in all comparison groups). The study also strives to address several methodological issues mentioned by previous literature regarding media-comparison studies (cf. All, Nunez Castellar, & Van Looy, 2016; Clark, 2012; Mayer, 2014a): proper randomization, the teacher effect, the effect of using different learning materials in different comparison groups, and the length of exposure effect.

--- Insert Table 1 around here ---

2. Study Background

2.1. Generalized Positive Valence Affect and Flow State

Various constructs, lying between emotions, motivation, and attention, have been used to study the impact of DGBL: including interest, engagement, intrinsic motivation, positive emotions, and flow. For the sake of simplicity, we denote all of these variables as *affective* variables (or affective states).

Despite the profound influence of affective states on higher level cognition (e.g, Blanchette & Richards, 2010; Isen, 2001; Linnenbrink & Pintrich, 2004) and on memory (e.g., Reisberg & Heuer, 2004; Linnenbrink & Pintrich, 2004), few research projects have investigated the influence of affectivity on learning in the context of technology-based instruction; especially in the context of DGBL (see references in Section 1) and the closely-related field of multimedia learning (e.g., Leutner, 2014; Park, Knörzer, Plass, & Brünker, 2015). Much is expected from the role of positive affective states in enhancing learning in the DGBL context, but little is known (which motivates our Research Question (c)).

In this study, we will employ two affective constructs: flow (Csikszentmihalyi, 1975) and generalized positive valence affect (Watson, Clark, & Tellegen, 1988). Flow is usually defined as pleasant absorption by an activity one undergoes (Csikszentmihalyi, 1975). As such, it is connected to increased attention to the object of the activity. Affect has a complex structure, but generalized positive and negative affect emerge as “two dominant and relatively independent [affect] dimensions” (Watson, Clark, & Tellegen, 1988, p. 1063). We will be interested here in the positive dimension. Various positive activating feelings, such as being enthusiastic, interested, alert, attentive, etc. (see Watson, Clark, & Tellegen, 1988), are associated with generalized positive affect.

These two constructs are complementary: flow is more related to attentional processes, while generalized positive affect relates to positive feelings. Together they can indicate if the learner is *positively activated* (cf. Pekrun & Linnenbrink-Garcia, 2012) when undergoing the instructional activity.

Various other overlapping constructs are related to "positive activation", such as intrinsic motivation (Ryan & Deci, 2000) or situational interest (Hidi & Renninger, 2006). These constructs are not the same, but differences between them are not important for our present purposes. Their levels also tend to be highly correlated in intervention studies (e.g., Brom et al., 2014a; Plass et al., 2013).

To address our primary Research Question (c), we first need to know if our game-based media induce comparably higher positive affect and flow (i.e., the answer to the first part of Research Question (a)). Three points justify the idea that this could be so. First, there is empirical evidence that positive activation tends to be experienced often when participants learn by advanced learning technologies, including games (D'Mello, 2013). Second, there is general agreement on the motivational aspects of games (e.g., Malone, 1981; Garris, Ahlers & Driskell, 2002). Third, the *Europe 2045* game has already been successfully used in Czech high schools.

2.2. Competitiveness and Positive Affect/Flow

Ambiguous results have been reported regarding the affective and cognitive advantages of competitive elements in the DGBL (Ke, 2008; Plass et al., 2013; Vandercruysse et al., 2013; see also ter Vrugte et al., 2015; DeLeeuw & Mayer, 2011). However, the forms of competition assessed by these studies were diverse and unlike *Europe 2045*'s competition. They typically involved either competition in dyads, competition by a single player against a virtual opponent or against the remainder of the class, or competition to achieve a tangible prize. *Europe 2045* features competition in larger teams (of at least 6 players) and with facets thought to promote learning, as detailed later in this section and in Section 4.3.

From a general perspective, it is known that organizing classroom instruction around competition is cognitively less effective compared to collaborative organization (see Johnson, Maruyama, Johnson, Nelson, & Skon, 1981; Qin, Johnson, & Johnson, 1995), though not necessarily compared to individualistic efforts (Johnson et al., 1981). This is a finding that is accommodated in the Social Interdependence Theory (Johnson & Johnson, 1989). This theory stresses the beneficial effects of the interdependent goal structures of peer learners and their individual accountability on learning. Shared, interdependent goals (absent in typical competitive situations) lead to promotive interactions. Individual accountability, where the performance assessment of each learner is available both to the individual and to the peer learners, may strengthen feelings of personal responsibility for the whole group of learners.

Still, competition in general can be constructive in educational settings, as long as it has the following features (summarized by Johnson & Johnson, 2009, and also integrated into the Social Interdependence Theory: Johnson & Johnson, 1989): all participants have a reasonable chance of winning, the rules and criteria for winning are clearly specified, and competition is not intense (i.e., winning is relatively unimportant, there are no tangible rewards for winning and no consequences for/impacts on students' grades). These are the features of *Europe 2045*'s competition (and thus of our both game-based media). The game's competition also features some collaborative aspects, especially positive goal interdependence among some peers, and provides informative feedback, known to enhance learning (see Hattie & Timperley, 2007).

Therefore, despite ambiguous findings from DGBL literature regarding the benefits of competition, we have reason to believe that *Europe 2045*'s competition can be advantageous for learning, and thus – in terms of our Research Question (a) – that it will contribute to the educational effectiveness of our two game media. At the same time, the classroom goal structure created by *Europe 2045*, which combines a certain mild form of competition with a touch of collaboration, is also employed in other games; for instance, in certain types of social role-playing games (cf. Mochocki, 2013). Information about these games' learning effectiveness is scarce. This study can thus contribute to our understanding of these games' advantages for learning.

However, different students have different attitudes toward competitive situations: some may like them while others may not. This simple point has not been, to the best of our knowledge, explicitly acknowledged in DGBL research. For this reason, we pose Research Question (b) and investigate the influence of learners' competitiveness, as a dispositional trait, on the positive affect and flow levels induced by the game media. We employ a two-dimensional conceptualization of competitiveness by Houston et al. (2002) and Harris & Houston (2010): the dimensions are *enjoyment of competitive situations* and *contentiousness*. We expect that the more intense these two characteristics are, the higher the positive affect/flow should be in the two game conditions. At the same time, we expect no influence of the two competitiveness traits on positive affect/flow when participants learn from the non-game medium, which lacks competitive elements. For the participants with the lowest levels of competitiveness, it is quite possible that the affective variables will be *higher* in the non-game medium. In short, we expect that competitiveness will moderate the influence of educational media on positive affect and flow.

2.3. Positive Affect/Flow as Mediators of Learning Outcomes

To address our primary Research Question (c), it is important to consider how positive affect/flow may influence learning outcomes. We will put forward a rationale based on the Cognitive-Affective Theory of Learning with Media (CATLM; Moreno, 2005), which is an expansion of a theory widely used in the field of multimedia learning (i.e., Cognitive Theory of Multimedia Learning; Mayer, 2009). The CATLM posits that learning effectivity depends on the effectivity of selecting relevant information by the learner from the instructional message, its organization into a coherent mental model in the learner's working memory, and integration of this model with the learner's prior knowledge. Efficiency of these processes depends, among other aspects, on the cognitive capacity available for these processes. Now, there is the following trade-off: learners do not always exert their total cognitive capacity, e.g. because they are bored. It is the cognitive capacity *actually used* (Moreno, 2010) that is important. On the one hand, positively engaged/activated learners (i.e., with a higher positive affect and flow level) can use more of their available cognitive capacity for learning-relevant processing, which improves learning. On the other hand, some details of the educational materials may be irrelevant for learning, such as those aspects of the educational game that increase the positive affect/flow in the first place. Because the learners must still process these details, part of their cognitive capacity is "consumed" by learning-irrelevant processing, which compromises learning.

The elements of computerized materials boosting positive affect are thus beneficial for learning only if they help in recruiting more cognitive capacity than is spent for their processing (see Mayer, 2014b, Park et al., 2015 for more on this trade-off). Therefore, poorly designed educational games can increase positive affect/flow and still hinder learning; unlike well designed games, for which higher positive affect/flow can contribute to learning.

Concerning Research Question (c), we have reason to believe that the latter would be the case for *Europe 2045* (because of its educational success in Czech high schools), i.e., that positive affect/flow would positively mediate learning outcomes.

Positive affect/flow may influence not only processes needed for initial knowledge acquisition, such as attention, information processing, or retrieval of prior knowledge (e.g., Isen 2001; Linnenbrink & Pintrich, 2004), but it could also possibly lead to a slowing down of participants' forgetting (cf. Reisberg, 2006). This was actually indicated by early reviews of (oft non-digital) game-based learning research (Pierfy, 1977; Randel, Morris, Wetzel & Whitehill, 1992). Therefore, we have reason to believe that knowledge decline, i.e., the difference between scores from tests administered immediately after the intervention and one month later, would be lower for the two game media.

2.4. Computers as a Delivery Technology for Playing Games

Particular educational content can be delivered by multiple educational methods, such as self-reading, frontal lecture, or by various types of collaborative activities. These educational methods can be, in turn, conveyed through different instructional media, such as a non-game workshop, a computer game, or a non-computer game. Our Research Question (d) asks if there are differences relevant for learning between computerized gameplay and gameplay using pen and paper. To answer this question, it is necessary to separate the effect of mere computer usage (in the context of game-based learning) from the effects of game playing.

Owing to the technological affordances of computers, digital educational games often use various learning aid tools. Examples of these, in the case of *Europe 2045*, include online student forums, a hypertext encyclopedia, dynamically modelled events students have to react to or teacher's statistics. The use of these tools, which are, by definition, unavailable in the matching non-computer game, often implies subtle changes to the educational methods. For example, the use of online communication tools alters interaction patterns in the classroom (and these patterns usually constitute part of the educational method).

To answer Research Question (d), it is necessary to use, with both game media, not only the same game with the same learning content, but also the same educational method. Otherwise, if a between-group difference is found, it would not be clear whether to attribute the difference to the different educational method or to the different delivery technology (cf. All, Nunez Castellar, & Van Looy, 2016). For instance, Higgins, Mercier, Burd, and Joyce-Gibbons (2012) showed the advantages of undertaking a game-based learning activity using a multi-touch table compared to a paper-based version of this activity. However, the difference was probably caused by different affordances of the multi-touch technology (and not due to merely delivering the learning experience through a different technology). Students could enlarge or shrink digital slips of "paper" with crucial information in the multi-touch condition (but not in the paper-based condition), which encouraged joint attention in the multi-touch condition and thus changed interaction patterns, and thereby the educational method. Therefore, for the study's purpose, it is necessary to remove as many of the tools altering the educational method as possible from the computer game in order to equate the methods.

It is widely presumed that a mere change of delivery technology (e.g., computers vs. pen and paper) during the otherwise same educational experience should not cause much of a difference regarding the experience's instructional effectiveness (Clark, 2012; see also Cuban, 2001; Morrison, 1994; but also Tamim et al., 2011). However, what if, for instance, a particular technology is not well accepted in a particular context? For example, if student adoption of tablet devices in schools depends partly on teachers' attitudes towards use of this technology in schools (cf. Courtois et al., 2014), playing games on tablets may hinder learning when the teacher dislikes these devices (unlike playing similar games without tablets). Likewise, what if different presentational formats (e.g., a computer screen vs. a blackboard) impose different cognitive loads on different types of learners? Concerning Research Question (d), we have no expectations regarding differences between the two game media. There are reasons to believe that no difference would be found, but there are also arguments to the contrary. It would be useful to find out, because the issue of the relative (dis)advantages of educational computer vs. non-computer games is underexplored.

3. This Study

This study investigates, using between-subject design, the influence of three instructional media on positive affect, flow levels, and learning outcomes. These media are as follows: a) an educational social role-playing game with mild competition, *Europe 2045*, played on computers (Brom, Šisler, & Slavík, 2010) (EU-comp); b) *Europe 2045* played without computers (EU-no-comp); and c) a non-game classroom-based workshop (Class). In all three conditions, learners, predominantly high school students, learn about the topic of the EU's policy agenda and the European Union's political direction by means of a specific debate-based educational method.

Based on considerations explained in Section 2, we put forward four research hypotheses and one exploratory goal:

H1: Both game media will induce comparably higher positive affect and level of flow (see Figure 1).

H2: Competitiveness moderates the effect of media on positive affect and flow levels. Specifically, we presume (H2a) no relationship between competitiveness and positive affect/flow as concerns the non-game medium but a positive linear relationship for both the game media. We also presume (H2b) that for the least competitive participants, positive affect/flow will be comparably higher with the non-game medium (see Figure 2).

H3: Both game media will enhance learning (compared to the non-game medium); both in terms of higher learning gains and lower knowledge decline (Figure 1).

H4: Positive affect and level of flow will positively mediate the influence of educational media on learning outcomes (Figure 1).

E1: What is the difference between the two game media in terms of positive affect, level of flow and learning outcomes?

--- Insert Figure 1 around here ---

--- Insert Figure 2 around here ---

This study models an entire school day in a laboratory (Brom, Šisler, Buchtová, Klement, & Levčík, 2012) and uses a stratified randomization (with the stratum being class). Teachers rotate randomly in the conditions. The treatments last about 7 hours; including the introduction and questionnaire administration.

We administer brief knowledge pre-tests and larger immediate post-tests and one-month delayed post-tests. We also administer in situ measurements of flow and generalized positive affect. The main independent variables are instructional medium, participants' gender, school quality and the following two participant traits: enjoyment of competition and contentiousness. The main dependent variables are the scores from the knowledge tests, positive affect scores and flow levels.

4. Method

4.1. Participants

Our aim was to obtain a heterogeneous sample from the game's target audience, i.e. adolescent and college participants (to recruit people with different competitiveness traits and background knowledge). We recruited 14 high school groups from average and above average urban schools in the Czech Republic (*n* = 304; 138 males, 166 females; Mean age = 16.3, *SD* = 1.15) and two additional groups of college participants (mainly students of computer science or psychology) (*n* = 31; 21 males, 10 females; Mean age = 22.2, *SD* = 2.3).[1] Each high school group consisted of one class and the experiment was part of the students' regular education because the topic is tied to Czech national curricula. We recruited classes whose teachers were willing to participate, and making sure to include diverse classes: both in terms of their quality as well as their subject specialization. In one college group, students participated for course credit; in the second one, they received 400 CZK (~ 20 USD) as compensation. We also recruited 60 students whose sole task was to just complete the tests; without undergoing any treatment (i.e., naive participants; see Sec. 4.3).

[1] We point out that we assessed participants' salivary cortisol in seven of these groups (*n* = 127). That is because cortisol levels are known to correlate with physiological arousal. This part of the study is irrelevant for present purposes, but we want to emphasize partial overlap in the dataset with a different study (Brom, Buchtová, Šisler, Děchtěrenko, Palme, & Glenk, 2014b), with a total sample size N = 171. The current study and the second study present, to a large extent, different (but parallel) data.

4.2. Questionnaires and Tests

The purpose of the *pre-questionnaire* was to solicit information about participants' gender and age and to gather information about their prior knowledge. To avoid cuing what should be remembered, this brief "pretest" was different from the knowledge tests administered after the intervention. It focused on general knowledge about the EU, rather than on specific knowledge taught by the intervention. We used five self-assessment questions and four knowledge questions (see Appendix A). Each question was assigned between 1-4, 1-5 or 0/4 points; giving us a possible score in the range of 5-38.

To measure participants' flow state during the treatment, we administered the *Flow Short Scale* (Rheinberg et al., 2003; see also Engeser & Rheinberg, 2008). In this study, we report the data from its first subscale measuring components of flow with ten 7-point Likert items. Flow questionnaires were analyzed using T-norms provided with the standardized Flow Short Scale (Rheinberg, 2004) (Cronbach $\alpha = .85$; possible score transformed via T-norms: 21-74). This questionnaire also contained one question on subjectively perceived difficulty, which is a construct thought to be related to germane cognitive load by some researchers (DeLeeuw and Mayer, 2008, but see also de Jong, 2010): "In comparison to other educational activities you usually participate in, this one is:" (9-point Likert item with a scale ranging from "easy" to "difficult").

To obtain information about participants' affective state during the treatment, we administered the *PANAS* (Positive and Negative Affect Schedule; Watson et al., 1988), which consists of two mood scales: one for positive and the other for negative affect. Each scale consists of ten 5-point Likert items (possible score: 10-50). In this work, we are interested only in the positive scale, denoted here as *panas+* ($\alpha = .87$).

After the intervention ended, participants filled in a *post-questionnaire*, from which only one question is relevant for the purpose of this study. This question asked about frequency of game-playing, and it had a scale 1-4 (1 – *less than 1 hour a week or never*; 4 – *more than 10 hours a week*).

As concerns knowledge acquisition, we tested only knowledge that can be acquired during the debates embedded in the treatments. Participants received four knowledge tests. As detailed later on in the text, each participant represented one "project" (i.e., a European political vision) and argued, during the experiment, for two policy changes at the European level. The four tests evaluated:

a) knowledge about the participant's own project and its relation to the projects of other learners in the given group;

b) knowledge about the content of one of the two policies for which the participant argued during the intervention;

c) knowledge about the process of negotiations on policy changes;

d) the names of all policies discussed that day (around 16 out of 32 possible policies).

Points (a) - (c) relate to conceptual and high-level skill memory. Point (d) relates, to some extent, to episodic memory. As concerns Points (a) and (b), each participant was tested based on his/her own project/policy.

The total test score was a sum of scores of these four tests (0-52 points[2]). The score from the immediate knowledge test will be denoted as *score1* and from the delayed test as *score2*.

The tests used a mixture of multiple-choice, short answer and open-ended questions and mental map drawings (see Appendix A). The open-ended and mental map questions were graded by two independent evaluators. Cohen's weighted kappa (Cohen, 1968) was in the range .68 - .91 for all questions, which we consider appropriate agreement.

After completing the knowledge tests, participants also filled in a short version of the *SIAS*, social interaction anxiety scale (Kupper & Denollet, 2011). This inventory is irrelevant for the present study.

A month after the intervention, participants filled in a second battery of knowledge tests (i.e., *delayed tests*). The questions were the same as in immediate tests, with just the order of the questions changed. Knowledge *decline* is computed as follows: score1 - score2. Participants also filled in several additional inventories. Only one of these is relevant for present purposes: the *RCI*, Revised Competitiveness Index (Harris & Houston, 2010). This instrument features 14 items with a 5-point Likert scale that can be divided into two subscales: *enjoyment of competition* (*RCI.comp*; nine items; $\alpha = .94$) and *contentiousness* (*RCI.cont*; five items; $\alpha = .79$). Note that this inventory seems to assess competitiveness as a stable trait (Harris & Houston, 2010).

---

[2] We created the tests and calibrated them on a sample different from the experimental sample. After the experiment, two additional questions had to be removed because there was no difference between experimental participants' and naive participants' scores from these two questions. The test score range is given after these two questions' removal.

4.3. Procedure and Interventions

We organized 16 different experimental days; one day for one participant group. The course of every day evolved according to a fixed "optimal" schedule (Fig. 3, 5) and the research team followed the schedule as closely as possible. The experiment took place at an experimental location outside schools. It started around 8 a.m. All teachers participating in the experiment were members of the research team.

After the introduction, participants filled in the pre-questionnaires. Then the participants received an introductory lecture about the EU (approx. 20 minutes, using PowerPoint slides). Three different persons rotated in and out of the teacher role.

The group was divided randomly into two or three subgroups after the lecture (based on the number of participants). Each subgroup was assigned one of the three media. In case there were two groups, the Class medium was always assigned. Then we chose randomly whether the second medium would be EU-comp or EU-no-comp. Participants were matched based on their pre-test scores. We also took care to achieve similar male/female ratios for every group in its subgroups (see Appendix B for details).

Each subgroup moved into a different room. The participants were instructed to avoid any interaction with other subgroups' participants until the experiment ended. Each participant was provided a pen and blank paper. In the EU-comp subgroup, each participant sat at a separate computer.

Each subgroup had its own teacher. We used a pool of eight teachers: all males younger than 35 years of age, with similar clothing style, short hair and similar speech and teaching styles. These teachers were randomly assigned to their positions. Each teacher had an assistant, who administered the questionnaires and helped with technical issues.[3]

After splitting into subgroups, each subgroup continued as described later, until the treatment interaction ended shortly after noon. Then, following a short break, we administered post-questionnaires and a battery of knowledge tests. Each subgroup was tested in its own room.

About a month later, we entered the school to administer subsequent knowledge tests and a few inventories, including the RCI. Students were not informed in advance. The testing period lasted 90 minutes. Students present in the delayed testing session, but not attending the original experiment, were also given the tests. These students were considered naive participants.

--- Insert Figure 3 about here ---

4.3.1. EU-comp Medium

This condition featured the computer version of the *Europe 2045* educational game. One possible way to play the game in schools is to make it part of a "project-day." In this

[3] Each subgroup also had one independent research observer, who coded students' verbal and non-verbal behavior during the discussions. These data are irrelevant for present purposes, but we want to emphasize the presence of another person in the room. We also point out that we assessed participants' salivary cortisol in seven groups four times during the experiment (see Footnote (1)).

study, we modeled such a “project-day” in a controlled laboratory environment. After forming the subgroups, students played the game for about five hours.

To equate the computer and the non-computer versions of the game as much as possible (see Section 2.4), we removed several in-game tools that exploit the affordances of computers (such as online student forums or dynamically modelled events). We also standardized the game (and thus put some additional constraints on it), so as to make the different courses of game-play comparable.

Students play *Europe 2045* in teams, while the teacher assumes the role of a moderator. Each student represents a member-state of the European Union. At the beginning, the game situation closely resembles the real state of affairs in today’s Europe. The game proceeds in rounds with each round representing one year.

In this study, the game was played by exactly eight players in six rounds. Students played two layers of *Europe 2045’s* game play: the economic layer and the diplomatic layer. In the economic layer, each student defines the domestic policy of his/her state, such as tax levels and the level of environmental protection (Fig. 4). We did not test students on knowledge acquired from the economic layer.

The diplomatic layer, which is most important for our present purpose, implements a debate-based educational method, which is a derivation of an educationally successful (Johnson, Johnson, & Smith, 1996) method called academic controversies (see Online Resource 1 for details). In this 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. A teacher moderates discussions on these changes. The discussions simulate negotiations in a wide array of EU institutions. We will now outline these in further detail.

--- Insert Figure 4 about here ---

--- Insert Figure 5 about here ---

Each player has his/her own project to try to push through at the European level. A project is a vision of how the EU should look in the future and it is formally defined by: a) a set of policies that should be put in place, b) a set that should be suspended, and c) a set to which the project is indifferent (e.g., the Green Europe project supports environmental protection and investment into alternative energy resources, while the Conservative Europe project strives to preserve traditional values). 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 particular policy change. There is thus a certain amount of positive, but also negative, interdependence among students' projects. The final appearance of Europe at the end of each game session is the result of intense negotiations and voting in a given player group. In this study, the game offered eight different projects; one for each student. Every project offered exactly four policies.

The intervention proceeded as follows (the other treatments differed in some points):

1. General framing: In the first two tutorial rounds, the teacher familiarized the players with the game's rules, and with controlling the game's user interface. He informed

players that they would compete against each other in order to win; but also informed them that they would need to collaborate with some of their peers to win.[4]

2. Project selection and role-playing: The players had three minutes for reading brief textual descriptions of all eight projects. They were then assigned the projects based on their preferences. Based on the project, the teacher assigned each player a member state to play and gave him/her a flag badge and a small flag stand so as to better identify with his/her state.
3. Project introduction: The teacher gave each player the textual description of his/her project (about two sheets of A4 paper) and its policies (4 x 1-2 sheets of A4 paper). The players then had exactly eight minutes to read their project description. Then they each had exactly one minute to present the project's main visions to their fellow players (Fig. 5).

In each of the subsequent four rounds (rounds 3 to 6), the following activities took place:

4. Playing the economic layer: Players were able to briefly control their states.
5. Policy selection: Exactly four players (selected by a computer) proposed a draft for a policy change.
6. Policy presentation: Each of these four players had exactly eight minutes to read expository texts about his/her proposed policy. Meanwhile the other four players could control their state or read materials about policies associated with their own projects or about policies proposed by the other four players. After the eight minutes

[4] The ranking, stemming from the students' performance on the diplomatic layer, served primarily as feedback. It also informed students *why* they hold a particular rank (i.e., what policies compatible with the student's project had been accepted). The ranking had no consequences for students' grades and no tangible reward was given for winning as part of the game.

had passed, a round of debates started. Students moved away from the computers and presented their drafts for policy changes (1.5 minutes). Opponents or other proponents could then react/ask questions during a discussion moderated by the teacher (2-3 minutes for each draft).

7. Negotiation: After the four presentations, the negotiations for or against support of the proposed policy changes started (5 minutes; not moderated by the teacher); especially with the students representing the neutral stance toward one of the four issues. Students often stood up, created small clusters, secretly negotiated outside the classroom, etc.
8. Voting: The students voted on each draft presented (at computers). The teacher presented the results at the beginning of the next round, including the players' current rankings in the game.

Since every project had four policies associated with it and each student presented a policy draft exactly twice, each student had to choose exactly two out of four policies of his/her own free will.

Students could acquire the following knowledge by playing the diplomatic layer, which we later on tested: knowledge acquired by reading expository texts about one's own project and its policies, and by preparing for presentations; by reading expository texts on other projects, including associated policies; by observing players' behavior when playing the respective project roles and by listening to them; and by participating in the presentation of drafts for policy changes and in subsequent negotiations. Concerning policies, we tested knowledge of policy changes presented in either the 4th or the 5th rounds (each participant presented just one policy in these rounds).

After the 5th round, when the game became most heated, we administered Flow and PANAS questionnaires.

4.3.2. EU-no-comp Medium

This condition featured *Europe 2045*'s diplomatic layer played without computers. The voting system was implemented in the classroom using a ballot box (Fig. 6). It was impossible to replace the game's economic layer easily, so it was absent in this condition.

Due to the procedure of assignment to groups (see Appendix B), we set up this treatment for 6-8 players; each playing a different project. Except for the number of players, the debates were organized as in the EU-comp treatment (Points 5-8 from the description of the EU-comp treatment above). Other procedures (Points 1-3) were also very similar, with the following exception: the roughly 15 minutes usually spent by the EU-comp players controlling their state (Point 4; Fig. 4) included instead an extra break and a longer voting process (the votes had to be counted manually).

--- Insert Figure 6 about here ---

4.3.3. Class Medium

This condition modeled, in a laboratory, a half-day workshop on the topic of the European Union, as it would be implemented within a school, without using *Europe 2045*. We strove to design the project day so that learners could obtain maximum learning benefits (i.e., "the best possible" replacement). The debates about the European Union were a natural part of this project day. The elements of the EU-comp's condition were replaced as follows:

1. General framing: The Class students were told that we were investigating a *new* “discussion-based workshop” (to address the novelty effect). The words “game” and “competition” were avoided.

2. Project selection and role-playing: Each Class learner was paired with an EU-comp (or EU-no-comp) learner and was assigned the peer’s project. Thus the Class learners could not choose their own projects. The teacher never told the Class learners that they represented their projects/states, they did not receive flag badges/stands and the teacher told them “to study a project” rather than “to play a project role”. Therefore, the role-playing aspects were substantially suppressed. Nevertheless, the project and the associated policies were interconnected in the same way as in the game conditions. Exactly as in the game conditions, the Class learners were instructed to read short project descriptions for three minutes.

3. Project introduction: It was the same as in the other conditions.

4. Economic layer: It was absent.

5. Policy selection: The teacher assigned each Class learner a policy to study and to present based on what his/her peer had chosen in the EU-comp (or EU-no-comp) group (i.e., no choice was allowed).

6. Policy presentation: The Class learners had eight minutes to study the assigned policy and 1.5 minutes to introduce it (as in the game conditions; using the same expository texts). After each presentation, the teacher invited other students to express their opinions regarding whether the policy should be applied in the EU or not, when considering the context of “their” project. They could also ask questions. The teacher moderated the discussion (2-3 minutes per each policy).

7. Negotiation: It was replaced by a discussion started by the following instructions from the teacher: “Now please think about how the political tendency/view you read about today at the beginning of class [i.e., the project], is related to the policies that have just been presented. It can relate to them positively, neutrally or negatively.” The teacher then called upon students to express their opinions about at least a few policies, and encouraged them to discuss their opinions with their peers.

8. Voting: It was absent. The time allotted for voting (and playing the economic layer, Point 4) in the EU-comp condition was filled in by an unrelated short film about an EU topic. The teacher showed the film at the very end of the workshop (around 20 minutes). Students also had two short breaks in the middle.

Finally, the introduction to the game was replaced by an unrelated 40-minute-long frontal lecture and by an unrelated 20-minute-long, pen-and-paper “warm up” mini-game (both on the topics of the EU).

There are several technical issues worth commenting on. First, the knowledge we tested could be acquired neither from the game introduction nor from the voting *per se*; nor from the Class medium’s replacements of these elements.

Second, in both game treatments, four students prepared themselves for policy presentations, while the other four read materials about policies associated with their own projects or read materials about policies proposed by their peers (or played the economic layer of *Europe 2045* in the EU-comp condition). Our pilot study showed that this format did not work well for the Class medium. Because the other four players were not motivated by the game, they did not read the respective materials carefully and they tended to become bored and irritated.[5] Because we strived for the "best possible replacement", we had to replace four rounds of the EU-comp medium with two "rounds" in the Class medium condition. In both of these "rounds", each participant prepared him/herself for the presentations that directly followed (i.e. all participants still presented a policy twice during the day). The Flow and PANAS questionnaires were administered after the 2nd "round" of discussions.

Third, there could have been 6-10 students in each Class subgroup (and not exactly eight as in the EU-comp subgroups). A particular policy was nevertheless always assigned once to avoid double exposure.

4.4.Data analysis

We analyzed data with statistical program R.3.2.3 (R Core Team, 2016). Correlations were evaluated using Pearson correlation coefficient. Effect sizes for correlation were classified according to Cohen (1988) into small ($r \sim 0.1$), medium ($r \sim 0.3$) and large ($r \sim 0.5$).

The main effect of medium was estimated by linear mixed model with random effect of class and subgroup[6], controlling for pretest score, age, gender, and quality of school. Quality was coded as "university", "better high school", "worse high school." The mediation and moderation analysis was conducted based on recommendations of Baron and Kenny

[5] Similar situations would arise in a regular class: many Czech students considered the topic of the EU to be boring.

[6] We use the term *groups*, *class groups* or simply *classes* to refer to 16 participants' groups (i.e., 14 high school classes and 2 college groups). We use the term *conditions* or *medium* to refer to the three experimental conditions. We use the term *subgroup* to refer to a part of the class: to 6-10 participants who were assigned to one condition together after the class had been split.

(1986). Because we are unaware of a standard method to compute effect sizes with mixed effect models, we devised our own procedure: For categorical explanatory variables, the effect size was calculated as the ratio of the estimated contrast and the residual standard error. For numerical explanatory variables, we compared the conditional means of the response corresponding to the lower and upper quartile of the explanatory variable. That is, we calculated inter-quartile range of the explanatory variable in each class group and denoted by MIQR the *median* of these *inter-quartile ranges*: the effect size was then calculated as the product of the estimated regression coefficient and the MIQR divided by the residual standard error. In this way, we estimated the effect size comparing a somewhat lower (lower quartile) and somewhat higher (upper quartile) value of the explanatory variable assuming that the remaining explanatory variables and the class were fixed. Similarly to Cohen’s *d* (Cohen, 1988), we classify effect sizes of numerical explanatory variables into small (~ 0.2), medium (~ 0.5) and large (~ 0.8).

5. Results

5.1. Participants characteristics

Of 335 recruited participants, data for 325 were analyzed. Participants were excluded primarily for leaving early during the experiment due to, e.g., a medical appointment. Of the included participants, 105 were with partly missing data from the immediate testing session and 40 from the delayed testing session, either due to technical problems or due to omission (see Figure 7). Thirty-eight participants did not come to post-tests. The participants with partly missing data were excluded only from statistical tests/analyses for which the missing data would have been needed.

We compared participants' trait characteristics and pretest score variable across the three conditions using one-way ANOVA. As shown in Table 2, there were no significant differences, thus we can assume the conditions were sampled equally.

--- Insert Figure 7 about here ---

--- Insert Table 2 about here ---

5.2.Descriptive statistics

Table 3 shows the descriptive statistics for the key dependent variables. Table 4 reports correlations between all variables involved in the main analysis.

--- Insert Table 3 about here ---

--- Insert Table 4 about here ---

5.3.Positive Affect, Flow and Competitiveness

To investigate Hypothesis 1, we examined the effect of media (i.e., EU-comp, EU-no-comp, Class) on panas+ and flow using linear mixed model with two media dummy variables measuring effects of a game (i.e., GAME; the two game media vs. Class) and computer (i.e., COMP; the two games against each other), the random interaction between the class and medium (i.e., the random effect of a subgroup nested in a class), the random effect of class, and four covariates: pretest score, age, gender and quality (three levels):

(1) Y_{ij} (*panas+/flow*) = $\beta_0 + \beta_1$ *pretest*$_{ij}$ + β_2 (*gender*$_{ij}$ = female) + β_3 *age*$_{ij}$ + β_4 (*quality*$_i$ = worse) + β_5 (*quality*$_i$ = university) + β_6 (*condition*$_{ij}$ = GAME) + β_7 (*condition*$_{ij}$ = COMP) + $\eta_i + \xi_{subgroup} + \varepsilon_{ij}$,

where $\eta_i \sim N(0, \sigma_\eta^2)$ represents the random intercept in i^{th} class group, $\xi_{subgroup} \sim N(0, \sigma_\xi^2)$ is the random effect of the subgroup, and $\varepsilon_{ij} \sim N(0, \sigma^2)$ denotes the random error of j^{th} student in i^{th} class group.

The 4th and 5th column of Table 5 summarize the estimated linear mixed model for flow and panas+. These results showed a significant effect for the GAME variable, which means that the Class medium is associated with lower flow and positive affect compared to the two game media (medium to large effect sizes). No significant effect was found regarding the COMP variable, i.e., no difference between the two game media was found.

Thus, **Hypothesis 1** was supported: both game conditions induced higher positive affect and flow compared to the non-game medium, after correcting for pretest score, gender, age, and school quality.

--- Insert Table 5 about here ---

To investigate Hypothesis 2, we first tested if enjoyment of competition or contentiousness is a general predictor of panas+ or flow. To this end, we added either enjoyment of competition or contentiousness as an explanatory variable to the model of type (1). Because Hypothesis 2 presumes no relationship between competitiveness and positive affect/flow as concerns the non-game medium, but a positive linear relationship for both the game media, we also removed from the model the COMP variable (which means that in the second model we treated both game media together):

(2) $Y_{ij}\,(panas{+}/flow) = \beta_0 + \beta_1\, pretest_{ij} + \beta_2\,(gender_{ij} = \text{female}) + \beta_3\, age_{ij} + \beta_4\,(quality_i = \text{worse}) + \beta_5\,(quality_i = \text{university}) + \beta_6\,(condition_{ij} = \text{GAME}) + \beta_7\,(RCI.comp/RCI.cont) + \eta_i + \xi_{subgroup} + \varepsilon_{ij}$.

Results showed a significant, but small, effect of enjoyment of competition on flow ($\beta_7 = 0.14$, s.e. = 0.06, effect size = 0.18, $p < .05$) and panas+ ($\beta_7 = 0.18$, s.e. = 0.05, effect size = 0.28, $p < .001$). The effect of contentiousness was not significant.

As the second step, we added interaction term into the model of type (2), i.e., "β_8 $(condition_{ij} = \text{GAME})\,(RCI.comp/RCI.cont)$". The interaction term β_8 was not significant in any of the four models.

This means that **Hypothesis 2a** was partly supported: there is a positive linear relationship between enjoyment of competition and panas+/flow, but in all three conditions (i.e., not only when the game media are considered). No relationship between contentiousness and panas+/flow was revealed (neither when the two game media were combined, nor when the non-game medium was considered). Because the non-game medium is associated with lower panas+/flow and the slope of the regression line between the posited moderators (RCI.comp, RCI.cont) and panas+/flow does not change significantly, **Hypothesis 2b** is not supported: we cannot conclude that the Class medium induced higher panas+/flow for the least competitive participants.

5.4.Learning Effects and their Relation to Positive Affect/Flow

In order to investigate Hypothesis 3, we examined the effect of medium on score1, score2 and decline using linear mixed model of type (1) with score1, score2 and decline as dependent variables (Table 5; Col. 1 – 3). The results showed effects for the Class condition, such that participants in this condition achieved significantly lower score1 (small to moderate effect size) and significantly lower score2 (moderate effect size) compared to the two game media. Their knowledge decline was marginally larger than knowledge decline of the game media participants (small effect size). No such effect was found regarding the COMP variable (i.e., when the two games were contrasted). Regarding differing contribution of our four knowledge tests to the differences between the Class and the other two media, the most influencing was the test on the names of all policies discussed that day, followed by the test on the process of negotiations, the test on the participant’s project and on the policy (see Online Resource 2). The Class medium was consistently worse than or equal to the two game media across all four tests in both time points.

Thus, **Hypothesis 3** was supported: the games improved learning and even slowed forgetting (after correcting for pretest score, gender, age and school quality). The effect sizes were in small to moderate ranges.

We now turn our attention to Hypothesis 4. The first two conditions (according to Baron & Kenny, 1986) needed to establish if positive affect/ flow positively mediate the influence of medium on learning outcomes (see Figure 1) already hold: the independent variable (i.e., educational medium) affect the potential mediator (i.e., panas+, flow) and the independent variable (educational medium) affect the dependent variable (score1, score2, decline). We now need to do the following: use the type (1) model with score1, score2 and decline as dependent variables and add to it panas+/flow as an independent variable, i.e. "β_8 (*panas+/flow*)". With the resulting six models, we need to inspect a) if panas+/flow affect score1/score2/decline, and b) if the effect of medium on score1/score2/decline is less than it is in the complementary models without the added β_8 parameter (which are models from Table 5, Col. 1 – 3).

The results concerning subquestion (a) are depicted in Table 6: the affective variables have a significant or marginally significant effect on learning variables, with the exception of the flow → decline combination. Effect sizes are much larger for panas+ compared to flow. As concerns subquestion (b), in the three models with panas+ as the explanatory variable (and score1, score2 and decline as dependent variables), the effects for the GAME variable (i.e., estimates of β_6) were not significant and their absolute sizes were around half of complementary effects from the models without the added β_8 parameter (from Table 5, Col. 1 - 3). This means that panas+ indeed mediates influence of educational medium on learning outcomes. In the three models with flow as the explanatory variable, the effect for the GAME variable remained significant or marginally significant and the absolute sizes of the estimates of β_6 decreased by less than 15%. This means that flow was not confirmed as the mediator.

--- Insert Figure 8 around here ---

Because panas+ and flow are related constructs, we also tested a model, in which both were used as explanatory variables at the same time, i.e.:

(3) Y_{ij} ($score1/score2/decline$) = $\beta_0 + \beta_1\ pretest_{ij} + \beta_2\ (gender_{ij} = \text{female}) + \beta_3\ age_{ij} + \beta_4\ (quality_i = \text{worse}) + \beta_5\ (quality_i = \text{university}) + \beta_6\ (condition_{ij} = \text{GAME}) + \beta_7\ (condition_{ij} = \text{COMP}) + \beta_8\ panas+_{ij} + \beta_9\ \text{flow}_{ij} + \eta_i + \xi_{subgroup} + \varepsilon_{ij}$.

The results attributed the power to explain the portion of the between-media differences in knowledge gain solely to panas+ rather than flow (Tab. 7).

Hypothesis 4 is thus supported as concerns panas+ but not as concerns flow. This is one of the key findings of this study. For illustrative purposes, the relationship between panas+ and delayed test score is depicted graphically in Figure 8.

--- Insert Table 6 about here ---

--- Insert Table 7 about here ---

5.5. The Differences between the Two Game Conditions and Perceived Difficulty

Concerning **Exploratory Goal 1**, the differences between the two game media were negligible in cognitive and affective dimensions (Tab. 6, β_7). This means that effect of computers as delivery devices (and the economic layer of the game absent in the EU-no-comp condition) is probably negligible.

There was also no difference between our three conditions in subjectively perceived difficulty (one-way ANOVA, $F(2, 310) = 0.46$, $MSE = 3.18$, $p = .63$, $\eta^2 < 0.01$). This indicates that intrinsic complexity of the interventions was probably similar in all the conditions (including the two games).

Concerning practical differences between the two game delivery technologies, our informal observation is that participants played the games very similarly. It took the teachers about 20% longer to explain controlling the computer game (including the economic layer), but the voting process took comparably longer when the game was played without computers. More preparation was involved with the EU-no-comp condition, but the advantage was avoiding possible technical issues with computers.

5.6. Other Results

We see that the school quality substantially influenced the resulting test scores (Tab. 6) such that students from worse high schools scored significantly lower (large effect sizes).

Positive affect and flow of these students was also lower (small to moderate effect sizes). These are meaningful outcomes, indicating that the knowledge tests were valid and the treatments were probably more useful (given higher panas+ and flow) for higher achieving class groups.

Differences between male and female participants were negligible concerning the key cognitive and affective variables (Tab. 6).

Both the correlation matrix (Table 5) and five linear mixed models of type (2) with frequency of game playing as explanatory variable (i.e., β_7) showed negligible influence of previous game playing experience on affective and cognitive variables (all ps > .1; negligible effect sizes). This indicates that *Europe 2045* can work even for non-gamers.

For the sake of completeness, we also explored the influence of two dimensions of competitiveness on learning gains using six linear mixed models of type (2) with enjoyment of competition or contentiousness as additional explanatory variables and score1, score2 or decline as dependent variables. This influence was negligible (all ps > .1; negligible effect sizes).

6. Discussion

We investigated a positive affect–learning link in the context of digital game-based learning. We postulated that if a specific debate-based educational method were framed within a game-based medium (specifically, within a social role-playing game with a mild form of competition and with some collaborative aspects), learners' generalized positive affect and flow states would be higher compared to delivering these debates via a non-game medium (Hypothesis 1). Likewise, we posited that learning would be enhanced with the game (Hypothesis 3). We also postulated that positive affect/flow state would mediate the influence of educational media on learning gains (Hypothesis 4). We also explored whether there would be any relevant differences (in terms of positive affect, flow, and learning outcomes) between the game being played on computers vs. without them (Exploratory Goal 1). Finally, because a form of competition was involved in the game, we also hypothesized that enjoyment of competition and contentiousness (i.e., participant traits) would be related to positive affect/flow in the two game conditions (Hypothesis 2). Despite the obvious nature of our research questions, this study is in fact one of the first to investigate them explicitly. Unlike many previous DGBL studies, we created very similar learning experiences in our three experimental conditions – doing our best to address specifically the teacher effect and the length of the exposure effect, while using the same learning materials.

6.1.Positive Affective States and Learning

Key findings indicated that *both* game-based conditions elicited a comparatively higher positive affect and flow (Sec. 5.3). Participants learned more when the educational debates were delivered via either of the games (Sec. 5.4), and positive affect, but not flow levels, positively mediated the influence of the educational medium on learning gains (Sec. 5.4). Thus, Hypotheses 1 and 3 were supported and Hypothesis 4 was partially supported. Because we found that the higher the positive affect was the lower memory decline was (Tab. 5, 6), there was influence not only between positive affect and cognitive processes involved in the initial knowledge acquisition, but also between positive affect and subsequent memory (cf. Reisberg and Heuer, 2004; p. 20).

This study thus uncovered one example of treatment that is able to instigate positive affect *and also* boost learning: a certain type of social role-playing game built round a specific form of debate. In terms of our explanatory framework, CATLM (Moreno, 2005), this means that our game-based treatments featured some elements increasing positive affectivity and these elements helped to recruit additional cognitive capacity for processing learning-relevant information: greater additional capacity than was spent for processing of these elements. The amount of distraction caused by these elements was probably low. In our opinion, these elements included mild competition with both negatively and positively interdependent learner goals (i.e., with a touch of collaboration), and team role-playing (see Table 1).

To what extent do our results generalize? Our finding runs, to some extent, parallel to several other findings from the DGBL field (Cordova & Lepper, 1996; Giannakos et al., 2013; see also Sabourin & Lester, 2014), the multimedia research field (Um et al., 2014; Plass et al., 2014; van Der Meij, 2013; Brom et al., 2014a) and motivation research (Vollmeyer & Rheinberg, 2006). Yet there are also studies in these fields with null results, mixed results, or negative results concerning the link between affective and cognitive variables (Adams et al., 2012; van Dijk, 2010; Iten & Petko, 2014; Plass et al., 2013; Ritterfeld et al., 2009; Stege et al., 2012). While there seem to be more positive findings than truly negative ones, it is clear that one cannot expect that results from one DGBL (or multimedia learning) study would automatically apply for a different treatment. What is important are the intervention elements, for whom these elements are used, and under what conditions. So far, studies have been diversified regarding participants' ages and their other characteristics; the type, instructional and entertainment quality of treatments; treatment elements assumed to influence affective variables; the context in which treatments were administered; the affective variables measured and the measurement instruments.

Therefore, the results of the current study can be, for now, probably generalized within the following context only. We can expect that similar competitive-collaborative, social role-playing games delivering debate-based educational methods (but diverse topics) and played by similar audiences over a similar amount of time will increase positive affect/flow and, consequently, learning gains: whether or not computers are used to play the game. This may seem to be a slightly limited finding, but then again, it is actually unlikely that the findings of any DGBL study would generalize beyond the genre of the game in question, its game mechanics/elements, the target audience and the way the game was used. We also note that such social role-playing games are used in classrooms, such as ones for teaching history (e.g., Mochocki, 2013), but information about their learning effectiveness is limited. Our finding thus not only provides new knowledge on an abstract level, it has practical relevance as well.

On a more theoretical level, above average generalized positive affect *plus* the experiencing of flow share many characteristics with a transient affective state called engaged concentration (Baker et al., 2010; cf. D'Mello and Graesser, 2012). It is thus possible that this study's outcome is an incarnation of a (hypothetical) general rule stemming from the CATLM: "higher engaged concentration → higher learning gains [provided the additional cognitive load is low]". However, to our knowledge, not only is precise operationalization of the engaged concentration presently lacking, this construct can actually be multidimensional. For instance, whereas in our case generalized positive affect was a stronger predictor of learning gains than flow state, the reverse seemed to be the case in the study by Brom et al. (2014a), where college participants learned individually how to brew beer during a two-hour educational simulation. The relationship between engaged concentration and learning can also be reciprocal and/or mediated, e.g., via changes in motivation (Pekrun and Linnenbrink-Garcia, 2012).

To conclude, systematic research is now needed to investigate the link between learning gains and various affective variables pertaining to positive affectivity. This should include precisely operationalizing the concept of engaged concentration. Parallel research questions asking what game elements are most beneficial for increasing positive affectivity and for learning are of equal importance.

6.2.Mild Forms of Competition

In this study, the enjoyment of competition (as a dispositional trait) was positively related to positive affect and the experiencing of flow, but contentiousness (as a dispositional trait) was not (Sec. 5.3). This, with one minor exception, applied to all three conditions. Less competitive participants did not enjoy the non-game medium more. Hypothesis 2a was thus only partially supported and Hypothesis 2b was not supported. At the same time, enjoyment of competition was not related to learning outcomes (Section **Chyba! Nenalezen zdroj odkazů.**).

These outcomes are probably caused by the fact that competition in *Europe 2045* featured collaborative aspects and also elements that were supposed to contribute to competition's constructiveness (summarized by Johnson & Johnson, 2009); winning was relatively unimportant, all participants had a reasonable chance of winning, and there were unambiguous rules and criteria for winning. Moreover, the "reward" arriving after every game's round (i.e., the current ranking of players, which included the list of already accepted policies compatible with the student's project) clearly provided participants informative feedback. Such "informative feedback" rewards tend to increase participants' intrinsic motivation (unlike tangible rewards, which tend to decrease it; see Deci et al., 1999; Cameron et al., 2001; see also Hattie and Timperley, 2007). If the positive effect of these elements was more pronounced for participants with lower levels of enjoyment of competition, that would explain the pattern in our findings. In any case, because types of competition detrimental to learning exist (Johnson et al., 1981), it is important to investigate what types of competition work for whom and in what kind of games.

Enjoyment of competition (as a dispositional trait) was also related to positive affect/flow in the non-game medium. More research into the relationship between enjoyment of competition (as a dispositional trait) and performance in the context of collaborative learning in general, and debate-based educational methods in particular, should help to elucidate this finding.

6.3.Team Role-Playing

The positive effect of role-playing has been documented in some areas (see McGregor, 1993), but it has not received as much attention as it deserves in the DGBL field. Non-computer social role-playing games are used in educational contexts (e.g., Gjedde, 2013; Mochocki, 2013), but information about their effectiveness is even more limited than in the case of digital games (see Bowman, 2014).

In this study, we cannot separate the effect of mild competition from the effect of team role-playing (in the two game conditions). Still, our finding is consistent with the idea that team role-playing contributed to an increase in positive affect and flow; and perhaps, in turn, in learning gains. In this regard, it is especially comforting that perceived difficulty was not higher in the game conditions compared to the non-game condition (Sec. 5.5), because there is some evidence that role-playing activities could present a burden for some learners. Specifically, in the case of *Europe 2045*, they can be stressful for social interaction-anxious male participants (Brom et al., 2014b). At the same time, team role-playing can be particularly effective when it gives learners a higher sense of control compared to the non-role-playing activity (see Pekrun, 2006), which was also the case of *Europe 2045*. Considering all points together, the results of this study provide justification for future research on the positive effects of various types of team roleplay in the DGBL context.

6.4. Computer Game or Non-Computer Game?

In this study, cognitive and affective outcomes were markedly similar in both game conditions (Section 5.5). For inevitable technical reasons, the two game conditions differed in three variables rather than just one (presence/absence of computers as a delivery technology; presence/absence of the game's economic layer; and average number of participants in one group (slightly lower in the non-computer game; see the description of the EU-no-comp treatment and Appendix B)). This is not ideal, but it is common in intervention studies that use complex authentic treatments (as opposed to artificial laboratory treatments). It is thus impossible to conclude with certainty that the presence of computers had no influence on learning. However, it seems most probable that the influence of all three variables was in fact small or negligible: both in affective and cognitive terms. At least, based on our informal observations, it seemed that the participants played both game variations similarly. Therefore, we now hold that the mere presence of computers neither enhances nor hinders learning in the DGBL context.

In light of decades of research on instructional technologies' impact on learning (see Clark, 2012; Cuban, 2001), the above result is not surprising. It is widely held that when potentially confounding variables are controlled for, it is reasonable to expect no educationally relevant differences when the same learning experience is delivered by two different technologies (Clark, 2012; Morrison, 1994). Still, asking if such a pattern will also hold in a new context (i.e., in DGBL) is justified. On the one hand, some DGBL proponents presume that *digital* media have certain hidden qualities compared to “older”, non-digital media as concerns learning (Prensky 2001; Tapscott, 1998); on the other hand, acceptance of digital games at schools is not automatically guaranteed (e.g., Bourgonjon, Valcke, Soetaert, & Schellens, 2010; Courtois et al., 2014). One could thus be uncertain as to whether playing games on computers in the formal schooling context enhances or hinders learning compared to playing non-computer games. This study supports the modest position that delivering game-based learning on computers probably is not a big deal for high school and college students.

That said, even if the lack of “computer effect” may not be surprising for educational researchers familiar with the history of instructional innovations in classrooms, it is relevant for those who study the instructional advantages of new affordances of computers. It suggests that if a computer game, or a similar application, is augmented with some tools that can be implemented only when using computers, the computerized version will outperform the non-computerized one: as long as the tools offer learning advantages (cf. Tamim et al., 2011; Higgins et al., 2012). (We point out here that such tools were purposefully removed from the computerized game in the current study.) Our finding is also important for those studying collaborative learning methods. By incorporating two game-type media into the study, we conducted what is called *media replication* (Ross & Morrison, 1989). Media replications are useful in suggesting the robustness of the impact of instructional methods on learning across more media. Our result demonstrated such robustness for the debate-based method embedded within a role playing, game-based activity featuring mild competition with a touch of collaboration. The method is similar to more widely used academic controversies (Johnson, Johnson, & Smith, 1996; Online Resource 1), which have been somewhat ignored by the game-based learning community. Our study thus indicates that academic controversies (and their derivations) are a promising educational method for team-based learning games; no matter what technology is used for playing the games.

6.5.Limitations

A recurring lament of many educational researchers is that media comparison studies are problematic due to a) many potentially confounding variables and b) the multi-dimensionality of the difference between the experimental and control treatments (see, e.g., Clark, 2012). The oft-mentioned remedy is to conduct value-added studies to isolate treatment elements that contribute most to learning (e.g., Vandercruysse et al., 2013). While we do agree with the gist of the criticism, we believe that the problem lies in individual studies rather than in the method per se. First, many confounding variables can be, at least to some extent, controlled for. Second, value-added studies are also subject to Type (b) criticism, because "elements" of interest are typically multidimensional constructs. For instance, as already stated, there are many types of competitions (e.g., cf. this study, Ke, 2008, and Plass et al., 2013). Which dimensions contribute to learning and which are detrimental to learning?

Rather than viewing one research method as being superior to another, we see them as complementary. Carefully conducted media comparison studies can suggest promising elements, which can be later investigated in detail using value-added studies (such as competition and role-playing). One should be prepared to tackle the confounding variables in both types of studies; likewise, one should be prepared for the fact that "elements" that once seemed elementary can later be deconstructed.

We did our best to equalize the learning experiences in the three conditions as much as possible: keeping the instructional medium as the only difference. However, it was not always possible to achieve this goal.

First, participants in both game conditions were engaged in four repetitions of the debates; each repetition with (around) four presentations. However, the participants having the non-game medium were engaged in two repetitions with (around) eight presentations each. It is possible that the latter format could be less effective, because acquiring a complex view of political directions and policies may need more repetitions. However, as stated (see Section 4.3.3), our pilot showed that the 2 x 8 arrangement worked better for the non-game condition than the 4 x 4 arrangement. Therefore, had we insisted on the 4 x 4 arrangement for the non-game condition, the game's positive effect would likely have been even more pronounced. This brings us to the following question for future research: is it the case that repeated debates must generally be implemented within games or engaging contexts, because they otherwise become boring due to a long exposure?

Second, based on the 2 x 8 arrangement, the non-game medium participants had 16 minutes for reading expository texts about policies, while the game media participants had the same 16 minutes *plus* another 16 minutes during which they could either read the expository texts, control the economy of their state (EU-comp medium only) or do nothing (see Fig. 5): some of them used that time to study. We encouraged Class participants to study the expository texts during longer breaks, but they rarely used the time for this. We have three reasons to believe that the effect of this extra time was small to negligible. A) In general, the higher the positive affect was, the greater the learning gains were. Would the positive affect of the non-game medium participants have increased had they been forced to study the expository texts longer? This seems unlikely, especially because they generally refused to study the texts during extra breaks – the texts per se were rather boring. B) The expository texts were about policies (and partly about projects) and there were small “game”–“non-game” differences regarding test questions on policies and projects (Sec. 5.4; Online Resource 2). C) The time for reading also differed somewhat between the two game conditions (because of the presence/absence of the game’s economic layer), but no notable differences in test scores were detected between the two games. It thus seems that there was enough time for reading in all the conditions. Rather than more time spent reading expository texts, it seems that the quality and the depth of policy presentations, discussions and subsequent negotiations (which were more heated and in-depth in game conditions) contributed to “game”–“non-game” differences.

Therefore, in our opinion, these two differences do not undermine the main research conclusions.

Retrospectively, another limitation of this study (and of many other DGBL studies with the level of competition as a manipulated variable) is that we did not measure the perceived level of competition. Note that we found that competitiveness, as a dispositional trait, had some influence on induced positive affectivity in the seemingly non-competitive treatment (i.e., the non-game condition). This could be due to either of the following two reasons. First, debates per se could be perceived as slightly competitive activities. Second, competitiveness (as a dispositional trait) is negatively related to social interaction anxiety (as a dispositional trait) in case of *Europe 2045* (Brom et al., 2014b) and social interaction anxiety could influence positive affectivity in treatments where participants have to interact with their peers. Had we measured the perceived level of competition, we could have better addressed this issue. Such a measure should be incorporated into future studies pertaining to competition.

Finally, it is possible that with a richer research method (e.g., videotaping the learning session), we could find some differences between computer and non-computer games that cannot be captured by written self-reports and knowledge tests. This could be an interesting research avenue.

References

Adams, D. M., Mayer, R. E., MacNamara, A., Koenig, A., & Wainess, R. (2012). Narrative games for learning: Testing the discovery and narrative hypotheses. *Journal of Educational Psychology, 104*(1), 235-249.

All, A., Nunez Castellar, E. P., & Van Looy, J. (2016). Assessing the effectiveness of digital game-based learning: Best practices. *Computers & Education*, *92*, 90-103

Baker, R. S. J. d., D'Mello, S. K., Rodrigo, M. M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive–affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies*, *68*(4), 223-241.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology*, 51(6), 1173-1182.

Blanchette, I., & Richards, A. (2010). The influence of affect on higher level cognition: A review of research on interpretation, judgement, decision making and reasoning. *Cognition & Emotion, 24*(4), 561-595.

Bowman, S. L. (2014). Educational Live Action Role-playing Games: A Secondary Literature Review. In S. L. Bowman (Ed.) *The Wyrd Con Companion Book 2014* (pp. 112-131): Los Angeles, CA: Wyrd Con.

Brom, C., Bromová, E., Děchtěrenko, F., Buchtová, M., & Pergel, M. (2014a). Personalized messages in a brewery educational simulation: Is the personalization principle less robust than previously thought? *Computers & Education*, *72*, 339-366.

Brom, C., Buchtová, M., Šisler, V., Děchtěrenko, F., Palme, R., & Glenk, L. M. (2014b). Flow, social interaction anxiety and salivary cortisol responses in serious games: A quasi-experimental study. *Computers & Education, 79*, 69-100.

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

Brom, C., Šisler, V., Buchtová, M., Klement, D., & Levčík, D. (2012). Turning high-schools into laboratories? lessons learnt from studies of instructional effectiveness of digital games in the curricular schooling system. *Lecture Notes in Computer Science: Vol. 7516. E-Learning and Games for Training, Education, Health and Sports* (pp. 41-53): Springer.

Cameron, J., Banko, K. M., & Pierce, W. D. (2001). Pervasive negative effects of rewards on intrinsic motivation: The myth continues. *The Behavior Analyst, 24*(1), 1-44.

Clark, R. E. (Ed.). (2012). *Learning from Media: Arguments, Analysis, and Evidence, Second Edition* (2nd. ed.): Information Age Publishing.

Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological bulletin, 70*(4), 213-220.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.)*: Hillsdale, NJ: Erlbaum.

Cordova, D. I., & Lepper, M. R. (1996). Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology, 88*(4), 715-730.

Csikszentmihalyi, M. (1975). *Beyond Boredom and Anxiety*: Jossey–Bass, San Francisco, CA.

Cuban, L. (2001). *Oversold and underused: Computers in the classroom*. Cambridge, MA: Harvard University Press.

Courtois, C., Montrieux, H., De Grove, F., Raes, A., De Marez, L., & Schellens, T. (2014). Student acceptance of tablet devices in secondary education: A three-wave longitudinal cross-lagged case study. *Computers in Human Behavior, 35*, 278-286.

De Jong, T. (2010). Cognitive load theory, educational research, and instructional design: some food for thought. *Instructional Science*, *38*(2), 105-134.

D'Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction, 22*(2), 145-157.

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological bulletin*, *125*(6), 627-668.

DeLeeuw, K. E., & Mayer, R. E. (2008). A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load. *Journal of Educational Psychology, 100*(1), 223-234.

DeLeeuw, K. E., & Mayer, R. E. (2011). Cognitive consequences of making computer-based learning activities more game-like. *Computers in Human Behavior, 27(5),* 2011-2016.

Elliot, A. J., & Pekrun, R. (2007). Emotion in the hierarchical model of approach-avoidance achievement motivation. In P. A. Schutz (Ed.), *Emotion in education* (pp. 57-73): Elsevier Academic Press.

Engeser, S., & Rheinberg, F. (2008). Flow, performance and moderators of challenge-skill balance. *Motivation and Emotion, 32*(3), 158-172.

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

Giannakos, M. N. (2013). Enjoy and learn with educational games: Examining factors affecting learning performance. *Computers & Education, 68,* 429-439.

Gjedde, L. (2013). Role Game Playing as a Platform for Creative and Collaborative Learning, *Proceedings of European Conference on Games Based Learning '13* (pp. 190-197).

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

Harris, P. B., & Houston, J. M. (2010). A reliability analysis of the revised competitiveness index. *Psychological reports*, *106*(3), 870-874.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research, 77*(1), 81-112.

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational psychologist*, *41*(2), 111-127.

Higgins, S., Mercier, E., Burd, L., & Joyce-Gibbons, A. (2012). Multi-touch tables and collaborative learning. *British Journal of Educational Technology*,*43*(6), 1041-1054.

Houston, J. M., Harris, P. B., McIntire, S., & Francis, D. (2002). Revising the Competitiveness Index. *Psychological Reports*, 90, 31-34.

Hussain, M. S., AlZoubi, O., Calvo, R. A., & D'Mello, S. K. (2011) Affect detection from multichannel physiology during learning sessions with AutoTutor. *Lecture Notes in Computer Science: Vol. 6738. Artificial Intelligence in Education* (pp. 131-138): Springer.

Isen, A. M. (2001). An influence of positive affect on decision making in complex situations: Theoretical issues with practical implications. *Journal of Consumer Psychology*, *11*(2), 75-85.

Iten, N., & Petko, D. (2014). Learning with serious games: is fun playing the game a predictor of learning success? *British Journal of Educational Technology*. (on-line ahead of print)

Johnson, D. W., & Johnson, R. (1989). *Cooperation and competition: Theory and research*. Edina, MN: Interaction Book Company.

Johnson, D. W., & Johnson, R. T. (2009). An educational psychology success story: Social interdependence theory and cooperative learning. *Educational researcher, 38*(5), 365-379.

Johnson, D. W., Johnson, R. T., & Smith, K. A. (1996). *Academic Controversy. Enriching College Instruction through Intellectual Conflict.*: ASHE-ERIC Higher Education Report Volume 25, No. 3. Washington, D. C: The George Washington University, Graduate School of Education and Human Development.

Johnson, D. W., Maruyama, G., Johnson, R., Nelson, D., & Skon, L. (1981). Effects of cooperative, competitive, and individualistic goal structures on achievement: A meta-analysis. *Psychological bulletin, 89*(1), 47-62.

Ke, F. (2008). Computer games application within alternative classroom goal structures: cognitive, metacognitive, and affective evaluation. *Educational Technology Research and Development, 56*, 539-556.

Kupper, N., & Denollet, J. (2012). Social anxiety in the general population: introducing abbreviated versions of SIAS and SPS. *Journal of affective disorders*, *136*, 90-98.

Leutner, D. (2014). Motivation and emotion as mediators in multimedia learning.*Learning and Instruction*, *29*, 174-175.

Linnenbrink, E. A., & Pintrich, P. R. (2004). Role of affect in cognitive processing in academic contexts. *Motivation, emotion, and cognition: Integrative perspectives on intellectual functioning and development* (pp. 57-87): Lawrence Erlbaum Associates.

Malone, T. W. (1981). Toward a theory of intrinsically motivating instruction. *Cognitive science, 4*, 333-369.

Mayer, R. (2009). *Multimedia Learning* (2nd ed.): Cambridge University Press.

Mayer, R. (2014a). *Computer Games for Learning: an Evidence-Based Approach*: The MIT Press

Mayer, R. (2014b). Incorporating motivation into multimedia learning. *Learning and Instruction*, 29, 171-173.

McGregor, J. (1993). Effectiveness of role playing and antiracist teaching in reducing student prejudice. *The Journal of Educational Research, 86*(4), 215-226.

Mochocki, M. (2013). Edu-Larp as Revision of Subject-Matter Knowledge. *The International Journal of Role-Playing*, 4, 55-75.

Moreno, R. (2005). Instructional technology: Promise and pitfalls. *Technology-based education: Bringing researchers and practitioners together* (pp. 1-19): Information Age Publishing.

Moreno, R. (2010). Cognitive load theory: More food for thought. *Instructional Science, 38*(2), 135-141.

Morrison, G. R. (1994). The media effects question:"Unresolvable" or asking the right question? *Educational Technology Research and Development*, *42*(2), 41-44.

Park, B., Knörzer, L., Plass, J. L., & Brünken, R. (2015). Emotional design and positive emotions in multimedia learning: An eyetracking study on the use of anthropomorphisms. *Computers & Education, 86,* 30-42.

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational psychology review*, 18(4), 315-341.

Pekrun, R., & Linnenbrink-Garcia, L. (2012). Academic Emotions and Student Engagement *Handbook of Research on Student Engagement* (pp. 259-282): Springer Science+Business Media.

Pierfy, D. A. (1977). Comparative simulation game research: Stumbling blocks and stepping stones. *Simulation & Games*, 8, 255-268.

Plass, J. L., Heidig, S., Hayward, E. O., Homer, B. D., & Um, E. (2014). Emotional design in multimedia learning: Effects of shape and color on affect and learning. *Learning and Instruction, 29,* 128-140. doi: 10.1016/j.learninstruc.2013.02.006

Plass, J. L., O'Keefe, P. A., Homer, B. D., Case, J., Hayward, E. O., Stein, M., & Perlin, K. (2013). The Impact of Individual, Competitive, and Collaborative Mathematics Game Play on Learning, Performance, and Motivation. *Journal of Educational Psychology, 105*(4), 1050-1066. doi: 10.1037/a0032688

Prensky, M. (2001). Digital natives, digital immigrants. *On the Horizon 9*(5), 1–6

Qin, Z., Johnson, D. W., & Johnson, R. T. (1995). Cooperative versus competitive efforts and problem solving. *Review of educational research, 65*(2), 129-143.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/ (Accessed 20 March 2016)

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

Reisberg, D., & Heuer, F. (2004). Memory for Emotional Events. In D. Reisberg & P. Hertel (Eds.), *Memory and Emotion* (pp. 3-41): Oxford University Press.

Reisberg, D. (2006). Memory for Emotional Episodes: The Strengths and Limits of Arousal-Based Accounts. *Memory and emotion: Interdisciplinary perspectives* (pp. 13-36): John Wiley & Sons.

Rheinberg, F., Vollmeyer, R., & Engeser, S. (2003). Die Erfassung des Flow-Erlebens [in German]. In J. Steinsmeier-Pelster & F. Rheinberg (Eds.), *Diagnostik von motivation und selbstkonzept* (pp. 261-279): Hogrefe.

Ritterfeld, U., Shen, C., Wang, H., Nocera, L., & Wong, W. L. (2009). Multimodality and interactivity: Connecting properties of serious games with educational outcomes. *Cyberpsychology & Behavior, 12*(6), 691-697.

Ross, S. M., & Morrison, G. R. (1989). In search of a happy medium in instructional technology research: Issues concerning external validity, media replications, and learner control. *Educational Technology Research and Development, 37*(1), 19-33.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary educational psychology, 25*(1), 54-67.

Sabourin, J., & Lester, J. (2014). Affect and Engagement in Game-Based Learning Environments. *IEEE Transactions on Affective Computing*, 5(1), 45-56.

Sitzmann, T. (2011). A meta-analytic examination of the instructional effectiveness of computer-based simulation games. *Personnel Psychology, 64*(2), 489-528.

Stege, L., Van Lankveld, G., & Spronck, P. (2012). Teaching high school physics with a serious game. *International Journal of Computer Science in Sports*, 11(1).

Tamim, R. M., Bernard, R. M., Borokhovski, E., Abrami, P. C., & Schmid, R. F. (2011). What forty years of research says about the impact of technology on learning a second-order meta-analysis and validation study. *Review of educational research, 81*(1), 4-28.

Tapscott, D. (1998). *Growing up digital: the rise of the Net generation.* New York: McGraw-Hill.

ter Vrugte, J., de Jong, T., Vandercruysse, S., Wouters, P., van Oostendorp, H., & Elen, J. (2015). How Competition and Heterogeneous Collaboration Interact in Prevocational Game-Based Mathematics Education. *Computers & Education, 89*, 42–52.

Um, E. R., Plass, J. L., Hayward, E. O., & Homer, B. D. (2012). Emotional Design in Multimedia Learning. *Journal of Educational Psychology, 104*(2), 485-498. doi: 10.1037/a0026609

van der Meij, H. (2013). Motivating agents in software tutorials. *Computers in Human Behavior, 29*(3), 845-857.

van Dijk, V. (2010). *Learning the triage procedure. Serious gaming based on guided discovery learning versus studying worked examples.* (Master thesis), University Utrecht, the Netherlands.

Vandercruysse, S., Vandewaetere, M., Cornillie, F., & Clarebout, G. (2013). Competition and students' perceptions in a game-based language learning environment. *Educational Technology Research and Development, 61*(6), 927-950.

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

Vollmeyer, R., & Rheinberg, F. (2006). Motivational effects on self-regulated learning with different tasks. *Educational Psychology Review, 18*(3), 239-253.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology, 54*(6), 1063-1070.

Wouters, P., van Nimwegen, C., van Oostendorp, H., & van der Spek, E. D. (2013). A Meta-Analysis of the Cognitive and Motivational Effects of Serious Games. *Journal of Educational Psychology, 105*(2), 249-265. doi: 10.1037/a0031311

Figures

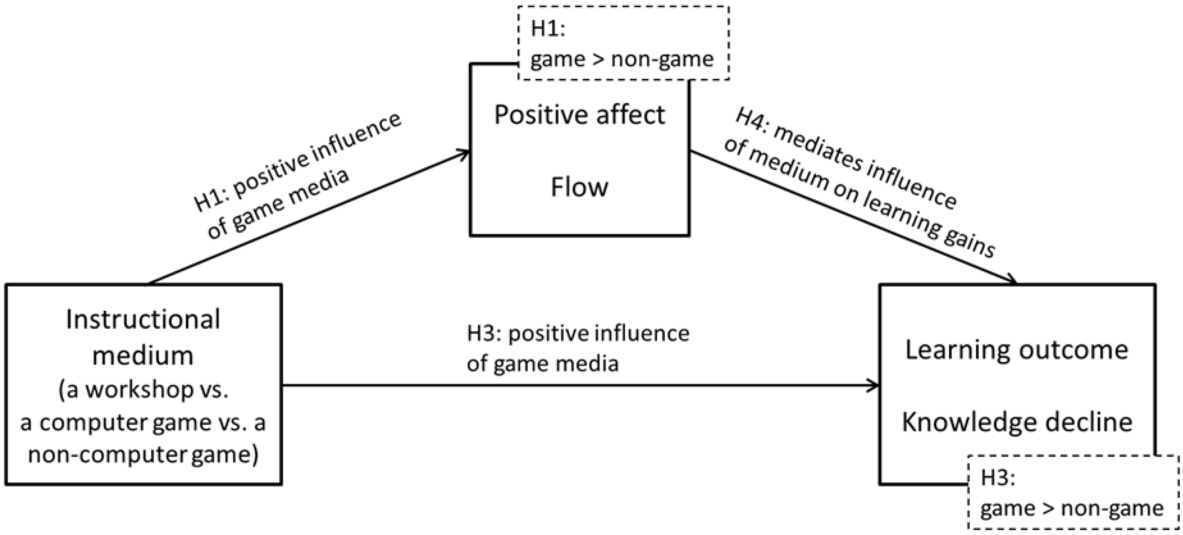

Figure 1. Schematic depiction of this study’s hypotheses H1, H3, and H4.

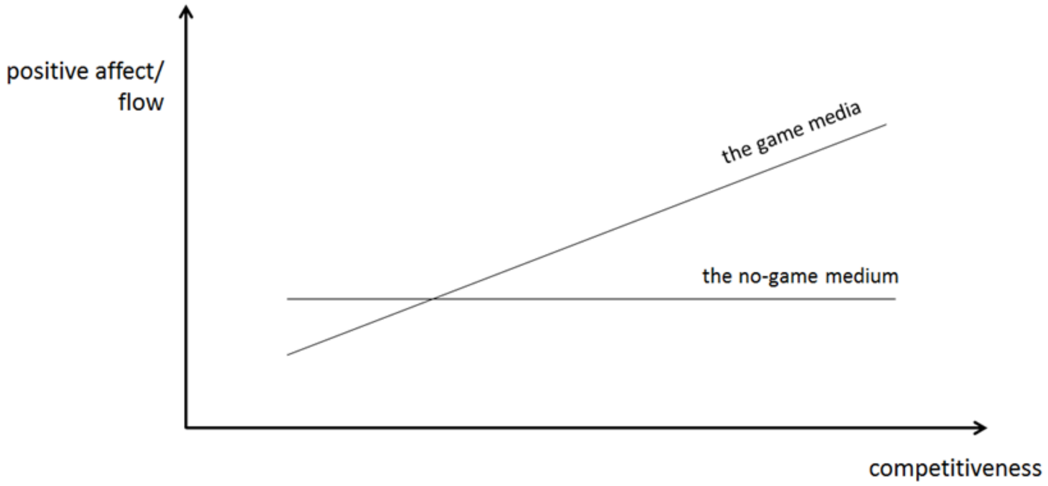

Figure 2. Schematic depiction of the hypothesized influence of competitiveness on affective variables (H2).

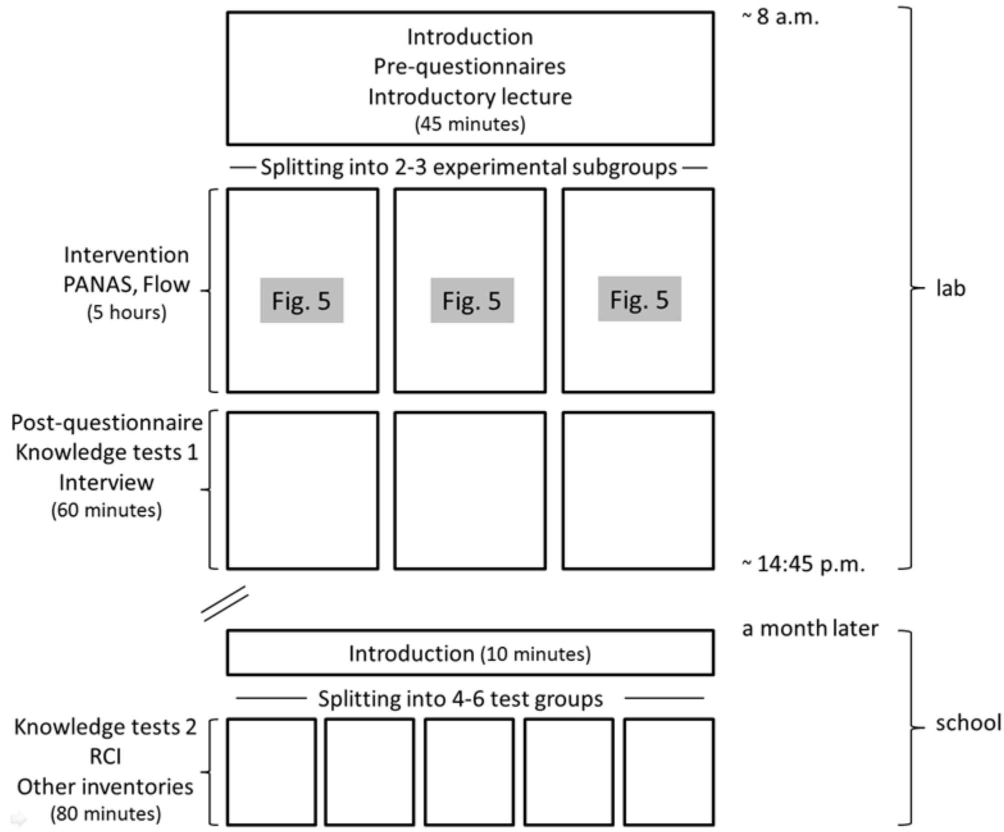

Figure 3. General schedule.

Figure 4. A screenshot from the *Europe 2045* game. The economic layer: GUI of domestic politics settings. Adopted from Brom, Šisler, & Slavík, 2010.

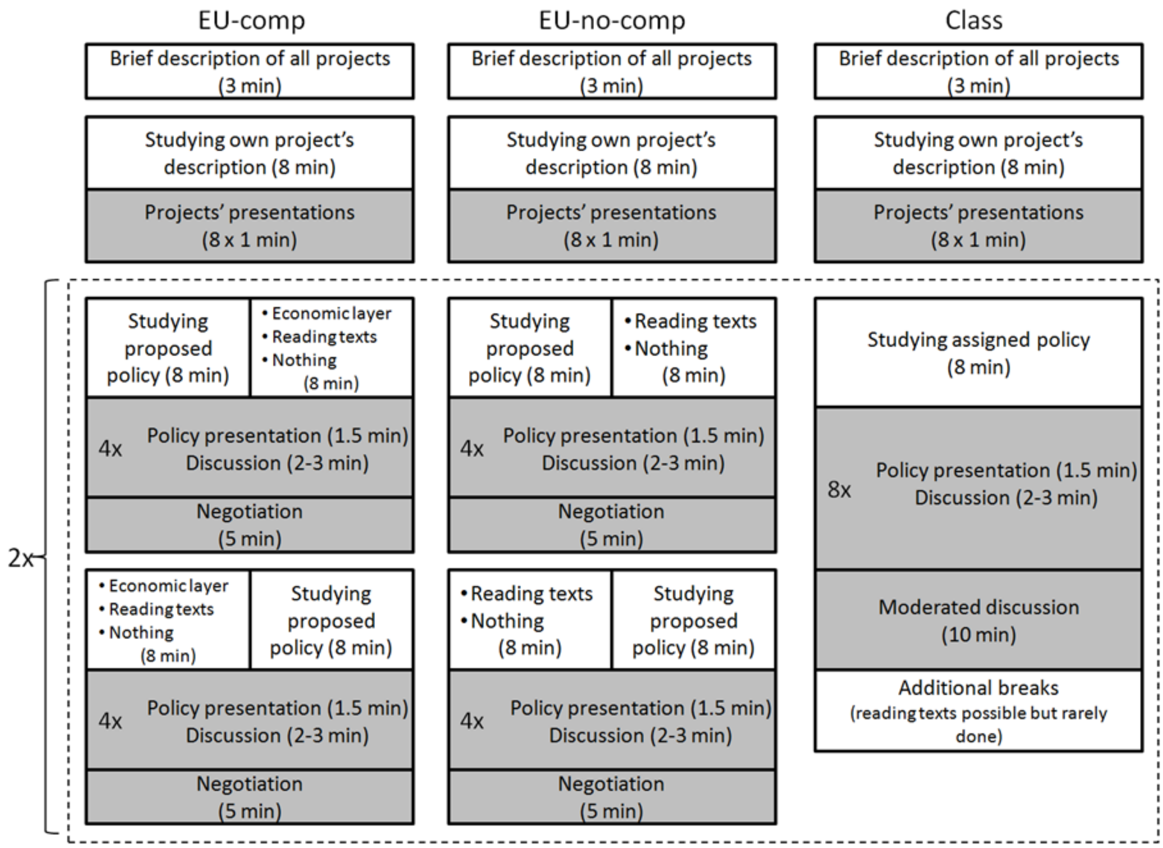

Figure 5. Comparison of the learning activities in the three conditions.

Figure 6. Voting using different media. Left: *Europe 2045* voting interface in the EU-comp treatment. Nine ballots for nine drafts of policy changes are depicted. Right: A teacher standing next to the ballot box announcing the results in the EU-no-comp treatment. The most recent proposals are written on the board behind him/her (top), as well as the players' latest rankings (bottom). There are ballots for individual drafts of policy changes on the table with the ballot box. Adopted from Brom et al., 2014b.

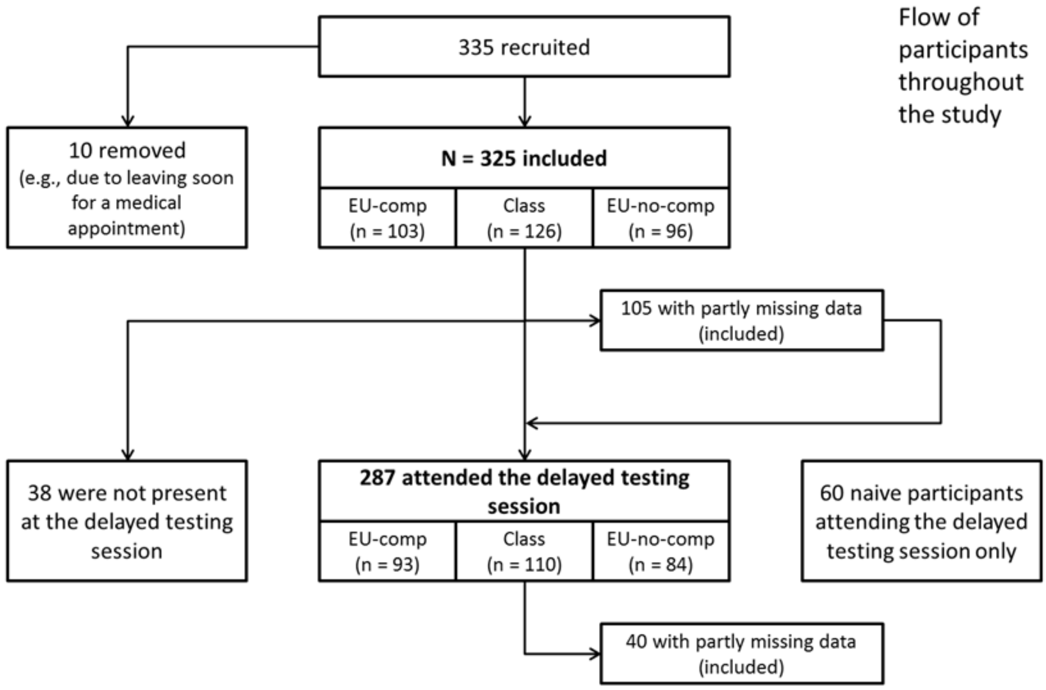

Figure 7. Sampling and flow of participants through the experiment.

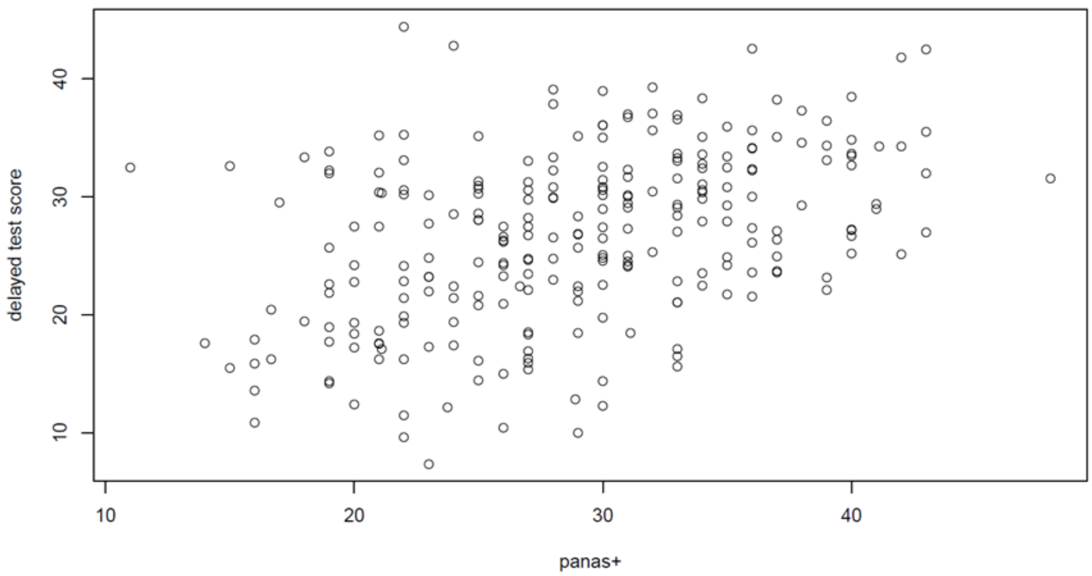

Figure 8. The relationship between one month delayed learning outcomes (total test score) and generalized positive affect.

Tables

Table 1

The crucial differences between the educational media used in this study.

	Medium		
	computer game	non-computer game	non-game workshop
Technology	computers	pen and paper	pen and paper
Media	mild competition	mild competition	-
features	team role-playing	team role-playing	-
Educational method	debates	debates	debates
Topic	EU's policy agenda	EU's policy agenda	EU's policy agenda

Table 2

Means and standard deviations (in parentheses) for participants' trait characteristics and pretest score for the three conditions

	Condition			
Variable	EU-comp	EU-no-comp	Class	*p*[a]
Pretest score	20.50 (5.32)	20.76 (5.47)	20.69 (4.83)	.94
Frequency of game playing	1.69 (1.04)	1.75 (1.00)	1.80 (1.01)	.61
RCI.comp	33.29 (7.37)	31.94 (7.56)	31.74 (8.64)	.39
RCI.cont	16.61 (3.96)	16.01 (4.25)	16.38 (4.67)	.53

[a]*p* values are for ANOVA model with random effect of class.

Table 3

Means and standard deviations (in parentheses) for the key dependent variables for the three conditions

	Condition		
Variable	EU-comp	EU-no-comp	Class
Score1	31.35 (5.77)	31.60 (5.29)	29.58 (7.58)
Score2	28.15 (6.06)	28.01 (6.28)	24.19 (8.51)
Decline[a]	2.84 (4.28)	3.51 (5.02)	4.58 (4.47)
Panas+	30.95 (6.34)	30.84 (7.21)	26.00 (6.77)
Panas–	17.86 (6.11)	18.00 (6.27)	18.06 (6.15)
Flow	50.86 (8.28)	49.65 (8.36)	46.18 (7.77)

Note. Some of these data were already presented in the study (Brom et al., 2014b) that has a partial overlap in the dataset with the present study. The data in this table somewhat differs from those presented in the previous study. The reason is that the present table concerns the whole sample, but the previous study only a subsample (n = 127).

[a]A higher value means a higher decline.

Table 4

Correlations between measures

Variable	2	3	4	5	6	7	8	9	10	11
1. Pretest	-.17**	.24***	.11†	.18**	.14*	.17**	-.01	.24***	-.01	.26***
2. SIAS	—	-.39***	-.43***	-.02	.08	.03	.04	-.10†	.25***	-.23***
3. RCI.comp		—	.30***	.12*	.12†	.14*	-.05	.29***	-.08	.24***
4. RCI.cont			—	.04	.05	.06	.03	.08	-.12*	.13*
5. Freq. game[a]				—	.02	.01	.08	.08	.08	.08
6. Score1					—	.78***	.21**	.26***	.10	.21***
7. Score2						—	-.45***	.42***	.14*	.30***
8. Decline							—	-.21**	-.06	-.07
9. Panas+								—	-.11*	.63***
10. Panas–									—	-.29***
11. Flow										—

Note. Some of these correlations were already presented in the study (Brom et al., 2014b) that has a partial overlap in the dataset with the present study. Correlations presented here somewhat differ from those presented there, because the present table concerns the whole sample, but the previous study only a subsample (n = 127).

[a]Frequency of game playing.

†$p < .10$ *$p < .05$ **$p < .01$ ***$p < .001$

Table 5

Estimates for models of five dependent variables

Explanatory variables	Dependent variables				
	Score1	Score2	Decline	Flow	Panas+
Pretest (β_1)	0.12 (0.07)	0.16 (0.08)	-0.02 (0.07)	0.37 (0.09)	0.27 (0.07)
	[0.13]	[0.17]*	[-0.02]	[0.31]***	[0.26]***
Gender (= F) (β_2)	0.12 (0.75)	0.56 (0.83)	-0.33 (0.66)	-1.29 (0.93)	-1.26 (0.79)
	[0.02]	[0.10]	[-0.07]	[-0.18]	[-0.20]
Age (β_3)	0.02 (0.31)	0.25 (0.41)	-0.20 (0.25)	0.40 (0.44)	0.05 (0.35)
	[0.00]	[0.05]	[-0.04]	[0.05]	[0.01]
Quality[a]	p = .002**	p = .006**	p = .991	p = .088†	p = .097†
Quality	-4.45 (1.07)	-5.21 (1.46)	-0.05 (0.70)	-3.32 (1.52)	-2.44 (1.13)
(= worse) (β_4)	[-0.83]**	[-0.93]**	[-0.01]	[-0.46]*	[-0.39]*
Quality	5.64 (2.27)	4.85 (3.00)	-0.20 (1.67)	-3.82 (3.29)	-2.56 (2.57)
(= univ.) (β_5)	[1.06]*	[0.87]	[-0.04]	[-0.53]	[-0.41]
Condition[a]	p = .033*	p = .000***	p = .099†	p = .000***	p = .000***
GAME (β_6)	1.77 (0.76)	3.36 (0.83)	-1.30 (0.64)	3.97 (0.90)	4.78 (0.73)
	[0.33]*	[0.60]***	[-0.29]†	[0.55]***	[0.76]***
COMP	-0.82 (1.01)	-0.11 (1.10)	-0.46 (0.82)	1.64 (1.18)	0.69 (0.96)
(β_7)	[-0.15]	[-0.02]	[-0.10]	[0.23]	[0.11]
$\hat{\sigma}_\eta$	1.097	1.958	0.000	2.103	1.426
$\hat{\sigma}_\xi$	0.937	1.079	0.000	0.895	0.001
$\hat{\sigma}$	5.278	5.522	4.497	7.198	6.255

Note. Standard errors are given in parentheses and effect sizes in brackets.

[a]Tests of significance of factor variables with three levels (likelihood ratio test).

†$p < .10$ *$p < .05$ **$p < .01$ ***$p < .001$

Table 6

Estimates for β_8 parameter for six different models with various combinations of dependent (columns) and explanatory (rows) variables.

Explanatory variables	Dependent variables		
	Score1	Score2	Decline
Panas+	0.17 (0.05)	0.33 (0.06)	-0.14 (0.05)
	[0.28]**	[0.55]***	[-0.28]**
Flow	0.08 (0.04)	0.10 (0.05)	-0.02 (0.04)
	[0.16]†	[0.18]†	[-0.04]

Note. Standard errors are given in parentheses and effect sizes in brackets

†$p < .10$ **$p < .01$ ***$p < .001$

Table 7

Estimates for β_8 and β_9 parameters for three different models of type (3).

Explanatory variables	Dependent variables		
	Score1	Score2	Decline
Panas+ (β_8)	0.16 (0.06)	0.36 (0.07)	-0.18 (0.06)
	[0.27]**	[0.61]***	[-0.36]**
Flow (β_9)	0.00 (0.05)	-0.05 (0.05)	0.06 (0.05)
	[0.01]	[-0.10]	[0.14]

Note. Standard errors are given in parentheses and effect sizes in brackets

**$p < .01$ ***$p < .001$

Appendixes

A. Questionnaires and Tests

This appendix introduces self-assessment and knowledge questions from the pre-questionnaire and three of the four knowledge tests used in the study.

A.1 Prior Knowledge Questions from Pre-questionnaire

1. I follow events on the international political scene:

a. not at all

b. once a week: *(select whatever options apply)* TV, online, radio, print media, other sources……………………..

c. 2-3 times a week *(select whatever options apply)* TV, online, radio, print media, other sources……………………..

d. daily *(select whatever options apply)* TV, online, radio, print media, other sources……………………..

2. Are you able to explain what the accession criteria are for a country wishing to join the EU? *(indicate your ability on a scale of 1 (not at all) - 5 (definitely yes))*

3. On topics related to the European Union I consider myself to be: *(select one answer)*

a. A beginner. I know a little about it.

b. Slightly advanced. I have average knowledge.

c. Advanced. I know quite a bit.

d. I don't know anything. I am not interested in this topic.

4. When I hear about political events in the EU, I can imagine what influences political decisions. *(indicate your ability on a scale of 1 (not at all) - 5 (definitely yes))*

5. Subject – The Basics of Social Science: *(select one answer)*

a. This is my favorite subject.

b. I find it generally interesting. I am often interested in the topics discussed.

c. I am not really interested. Most topics do not interest me.

d. It is my least favorite subject. I literally have a negative relationship to the subject.

6. Who is the current president of the European Commission? *(select one answer)*

a. Herman Van Rompuy

b. Catherine Margaret Ashton

c. Vladimír Špidla

d. José Manuel Durão Barroso

7. How many member-states does the EU currently have? *(select one answer)*

a. 12

b. 15

c. 27

d. 28

8. When did the Czech Republic join the EU? *(select one answer)*

a. 1998

b. 2001

c. 2003

d. 2004

9. Štefan Füle is the Czech Republic Commissioner for: *(select one answer)*

a. Employment, Social Affairs and Inclusion

b. Enlargement and European Neighbourhood Policy

c. Agriculture and Rural Development

d. Health and Consumer Policy

A.2 Policy Test

Sample Policy Test. This test is for “Immigration” policy.

1) Please list five words or combinations of words that in your opinion best describe the topic of the EU Common Immigration Policy that you read about today. Please give a detailed response, as in the following example.

> Example: Please list five words or combinations of words that in your opinion best describe the topic of the Kyoto Protocol that you read about today.
>
> 1. reduction of greenhouse gas emissions
> 2. international treaty
> 3. global warming
> 4. IPCC
> 5. USA hasn’t signed yet

2) Write down five main benefits that an EU Common Immigration Policy would have for member-states and for the EU in general (or for its residents). Please draw on the same positions that you presented during today’s seminar. Imagine that you are summarizing your main, factual arguments in favor of introducing this policy during a meeting of the Council of the European Union.

3) Were a Common Immigration Policy for all EU member-states to be introduced, it is to be expected that it will limit, on the part of immigrants, abuse of ………… Fill in the missing text.

4) What are the positive impacts of legal migration for EU member-states?

a. It will lead to an inflow of financial resources that immigrants bring with them.

b. It will help with business and cultural exchange between countries.

c. It will reduce the degree of extremist behavior in society.

d. It will help counter the negative consequences of the overall aging of the European population.

5) The FRONTEX Agency:

a. Handles EU asylum policy

b. Coordinates cooperation between the border control services of individual member-states

c. Ensures the functioning of the EU Coast Guard along the coast of the Mediterranean Sea

d. Ensures the inclusion (integration) of immigrants in EU member-states

A.3 Project Test

Sample Project Test. This test is for the “Liberalism” project.

1. Some of the terms shown below relate directly to the issue of liberalism. Circle those terms. For the terms that do not relate to the issue of liberalism, please cross them out. Do not do anything to the other terms (i.e. do not circle them, do not cross them out).

1. human rights
2. cultural identity
3. Milton Friedman
4. anti-totalitarian
5. individualism
6. personal ownership
7. revolutionary
8. Winston Churchill
9. collectivism
10. John M. Keynes

2. Please write inside the empty oval the name of the political movement that you received. In the space around it, fill in key terms that relate to this political movement.

A.4 Negotiation Test

All students received the same Negotiation Test.

1. Describe in several sentences what negotiating steps you would take in order to achieve the implementation of this policy. Do not give detailed arguments, only list the steps in the negotiations.

2. List five words or combinations of words that in your opinion best describe the weaknesses and inadequacies of the EU’s current decision-making process.

B. Assignment Conditions

The assignment to subgroups occurred as follows: the optimal number of participants in each subgroup was eight. Table B1 shows how large the subgroups were when a number of participants other than 16 or 24 arrived. Participants were matched based on their pre-test score in the following way: in cases of 19 or less participants, pairs and usually also a few singles were formed (see Table B1). Singles were selected randomly. In cases of 20 or more participants, trios and usually also a few pairs or singles were formed. Members of the pairs/trios were then assigned to the subgroups randomly. Singles were assigned according to the table. In case this random assignment resulted in a situation in which the male/female ratio in the subgroups differed and could be improved by an exchange, the researchers swapped members of one or two randomly chosen mixed-sex pairs/trios. Sometimes, one or two students had to leave before the experiment's end. In such cases the student was assigned to the Class condition.

Table B1

Assignment to conditions

Size of the Whole Group	Condition Name		
	EU-comp or EU-no-comp	Class	EU-no-comp
15	8	7	
16	8	8	
17	8	9	
18	8	10	
19	8	11	
20	8	6	6
21	8	7	6
22	8	8	6
23	8	7	8
24	8	8	8
25	8	9	8
26	8	10	8

<u>On-line Resource 1:</u> Academic Controversies and Repeated Tri-stance Debates

All three treatments used in this study are organized around a specific debate-based educational method, which is naturally embedded in certain team-based games. In fact, this method is a blend of classical debates and a successful (Johnson, Johnson, & Smith, 1996) educational method called *academic controversy* (or just *controversy* in the following text). We call this method *repeated tri-stance debate* and the purpose of this text is a) to explain why it is so and b) to detail advantages and disadvantages of this method in comparison to academic controversies and classical debates.

Academic Controversies

For clarity, we first introduce classical academic controversy. It is a small-group, collaborative teaching method that encourages students to represent one of the extreme positions in a controversial, bipolar issue; argue in support of it with peer learners who represent the opposite position; then switch their views and eventually reach a consensus within the whole group (Johnson et al., 1996). The switching of views and arriving at a consensus is usually absent in pure bi-polar debates, which are an inferior educational method to controversies (Johnson et al., 1996). Academic controversies and debates share certain characteristics important for learning but differ in others. The shared characteristics, as conceptualized by Johnson et al. (1996; p. 8), are:

a) categorizing and organizing information to derive conclusions;

b) presenting, advocating, elaborating positions and rationales;

c) being challenged by opposing views;

d) conceptual conflict and uncertainty about the correctness of one's own views;

e) epistemic curiosity.

However, only controversies enable:

f) perspective taking;

g) engaging learners in reconceptualization, synthesis and integration.

Repeated tri-stance debates

We now introduce similarities and differences between controversies and repeated tri-stance debates. Similarly to controversies, repeated tri-stance debates create situations, in which students represent certain views, argue in support of them and discuss them with students representing opposite or *neutral* views. Students representing a neutral view can also side with either of the extreme positions at their will. Repeated tri-stance debates constitute the core of game mechanics of a subtype of social games, which includes games such as *Diplomacy*[7] and *Europe 2045*. In these games, discussions among students (players) become negotiations in order to pursue the game's goal.

The first key distinction between controversies and repeated tri-stance debates is the presence of the neutral views in the latter method. "Tri-stance" in the method's name means the three possible positions: negative – neutral – positive. The second distinction is that the latter method enables repeated, thematically related, discussions (in games, these repetitions happens in different game *rounds*). This is reflected by the word "repeated" in the method's name. The third distinction is the absence of switching the views.

[7] *Diplomacy* is a strategic board game with intense negotiation phases and the near absence of random effects (for up to seven players) (see, e.g., Calhamer, 1974).

From the educational perspective, academic controversies and repeated tri-stance debates share the characteristics (a) – (f). As concerns the characteristics (g), there are some differences.

Controversies achieve Point (g) primarily via the switching of roles and reaching a final consensus. The repeated tri-stance debates do that by engaging learners representing opposite views in discussions with co-learners representing neutral views (in games, this can be followed by the whole group ultimately accepting, by majority vote, one of the bi-polar positions in order to proceed further in the game). This method may not be as good as controversies in achieving Point (g), which is a disadvantage. However, it also has one advantage due to its repetitive nature and the possibility to represent the neutral position. It allows for representation of mutual relationships among multiple issues such that a neutral stance toward one issue can be inherently connected to either a negative or positive stance toward another issue (or more issues). For example, in the *Europe 2045* game, students represent political visions that they try to push through at the European level; one student may represent "liberalism" and the other "environmentalism". Both students may support the legalization of "same-sex civil partnership" but have opposing views on the issue of "environmental tax" (as specified by the game's rules). At the same time, their stance toward the so-called "European Presidency" may be neutral; unlike that of the representative of "European federalism", who would argue in favor of the Presidency, and that of the representative of "Euroskepticism", who would argue against the Presidency. Because students can be engaged in several consecutive discussions, this intricate relationship among political issues and umbrella political visions can be gradually revealed to them. Students can consequently acquire a complex view of the overall situation, which is hard to achieve through a one-round academic controversy.

Goal Structures

From the perspective of goal structures created by the learning situations, the crucial distinction between debates and academic controversies lies in positive social interdependence, and shared goals between learners in particular; present in the controversies and absent in debates (Johnson & al., 1996; Johnson & Johnson, 2009). In academic controversies, learners share the same goals: to learn collaboratively and to acquire a consensual view on the discussed topic. Repeated tri-stance debates bring about partly shared goals (when learners represent the same position on a bi-polar issue) and partly disjunctive goals (when learners represent opposing views). Because of the disjunctive goals, the repeated tri-stance debates inherently feature higher levels of competition compared to controversies. Because of the explicitly shared goals, they feature higher levels of collaboration compared to pure debates. Unlike both controversies and debates, they enable players representing neutral positions to join a shared goal (by supporting one of the extreme positions) or detach from that goal (by starting to support the other extreme position or becoming neutral again).

Different Classroom Goal Structures as an Alternative Explanation of the Study's Main Findings?

Participants in our study learnt more under both game-based conditions compared to the non-game condition and exhibited a higher positive affect (Hypothesis 1 and 3). The "game"–"no-game" differences the study found in positive affect and learning gains are explicable also in terms of differing classroom goal structures (for different conditions). The key advantageous feature of collaborative learning settings is positive social interdependency among learners; including the existence of a shared learning goal (e.g., Johnson & Johnson, 2009). This is also explicitly verbalized by the Social Interdependence Theory (Johnson & Johnson, 1989). Unlike in academic controversies, there is no explicitly stated goal shared by all learners in repeated tri-stance debates. However, our both game media could implicitly create a situation in which players enjoying the game could adopt a tacit shared goal – to proceed forward in the game and to change the simulated Europe. It is not the goal to learn collectively "as much as possible" and to acquire a consensual view on the issues discussed, but it is a goal that nevertheless can drive learning. Such a goal was absent in the non-game medium.

In the main article, we argued that a team role playing and a mild competition with collaborative aspects were two features most probably responsible for the affective and cognitive effectivity of the *Europe 2045* game. Are differing goal structures (game vs. no-game) a third possible aspect? Yes and no. It truly could contribute to the between-treatment differences. However, the situation in which the shared goal could be adopted was afforded by the game. And the game created this situation via its core elements; the underlying game structure and – the light-weight team role playing and the mild competition. In other words, whereas a mild competition and a team role playing are distinct elements, the differing goal structures, if really existed during game playing, were created (in the present case) to some extent due to the competition and the role playing.

References

Calhamer, A. (1974). The Invention of Diplomacy. *Games & Puzzles, 21*.

Johnson, D. W., & Johnson, R. (1989). *Cooperation and competition: Theory and research.* Edina, MN: Interaction Book Company.

Johnson, D. W., Johnson, R. T., & Smith, K. A. (1996). *Academic Controversy. Enriching College Instruction through Intellectual Conflict.*: ASHE-ERIC Higher Education Report Volume 25, No. 3. Washington, D. C: The George Washington University, Graduate School of Education and Human Development.

Johnson, D. W., & Johnson, R. T. (2009). An educational psychology success story: Social interdependence theory and cooperative learning. *Educational researcher, 38*(5), 365-379.

<u>On-line Resource 2:</u> Impact of Individual Knowledge Tests

The primary knowledge gain variables investigated in this study were total test scores: score1 and score2. These variables consisted of scores from four independent tests (see also Appendix A):

a) knowledge about the participant's own project and its relation to the projects of other learners in the given group (Project Test);

b) knowledge about the content of one of the two policies for which the participant argued during the intervention (Negotiation Test);

c) knowledge about the process of negotiations on policy changes (Policy Test);

d) the names of all policies discussed that day (around 16 out of 32 possible policies) (Episodic memory test).

Do differences in achievement on these four tests between the three conditions, i.e. EU-comp, EU-no-comp and Class, match the between-condition differences in score1 and score2?

We examined the effect of medium on the scores of these four knowledge tests administered immediately after the treatment and a month later using eight (4 x 2) linear mixed models of type (1) (see the main paper, Sec. 5.3) with individual test scores as dependent variables. The results are summarized in Tab. OR2.1. For comparison, the table also shows results for score1 and score2 taken from Tab. 5.

Table OR2.1 – first part

Estimates for models of eight individual test scores, score1, and score2 as dependent variables.

Explanatory variables	Dependent variables				
	Project-score1	Negot.-score1	Policy-score1	Epmem-score1	Score1
Pretest (β_1)	0.02 (0.04)	0.02 (0.02)	0.07 (0.03)	0.02 (0.01)	0.12 (0.07)
	[0.03]	[0.06]	[0.14]†	[0.14]*	[0.13]
Gender (= F) (β_2)	0.34 (0.39)	-0.22 (0.21)	-0.23 (0.37)	0.24 (0.09)	0.12 (0.75)
	[0.12]	[-0.13]	[-0.08]	[0.33]**	[0.02]
Age (β_3)	-0.17 (0.16)	0.03 (0.08)	0.10 (0.14)	0.04 (0.04)	0.02 (0.31)
	[-0.06]	[0.02]	[0.03]	[0.05]	[0.00]
Quality[a]	*p*=.008**	*p*=.000***	*p*=.003**	*p*=.535	*p*=.002**
Quality (= worse) (β_4)	-1.65 (0.54)	-1.10 (0.24)	-1.68 (0.38)	-0.14 (0.14)	-4.45 (1.07)
	[-0.57]*	[-0.65]***	[-0.58]***	[-0.19]	[-0.83]**
Quality (= univ.) (β_5)	2.71 (1.19)	1.85 (0.58)	1.07 (0.96)	0.19 (0.33)	5.64 (2.27)
	[0.94]*	[1.10]**	[0.37]	[0.26]	[1.06]*
Condition[a]	*p* = .834	*p* = .001***	*p* = .797	*p* = .000***	*p* = .033*
GAME (β_6)	-0.02 (0.38)	0.77 (0.22)	0.24 (0.35)	0.90 (0.14)	1.77 (0.76)
	[-0.01]	[0.46]**	[0.08]	[1.24]***	[0.33]*
COMP (β_7)	-0.27 (0.50)	-0.24 (0.28)	-0.00 (0.45)	-0.34 (0.17)	-0.82 (1.01)
	[-0.09]	[-0.15]	[-0.00]	[-0.46]†	[-0.15]
$\hat{\sigma}_\eta$	0.578	0.000	0.000	0.000	1.097
$\hat{\sigma}_\xi$	0.382	0.238	0.000	0.335	0.937
$\hat{\sigma}$	2.874	1.666	2.908	0.672	5.278

This table continues on the next page.

Table OR2.1 – continuation

Explanatory variables	Dependent variables				
	Project-score2	Negot.-score2	Policy-score2	Epmem-score2	Score2
Pretest (β_1)	0.04 (0.03)	0.03 (0.02)	0.06 (0.04)	-0.00 (0.01)	0.16 (0.08)
	[0.10]	[0.10]	[0.12]	[-0.03]	[0.17]*
Gender (= F) (β_2)	0.22 (0.37)	0.27 (0.24)	0.10 (0.42)	0.15 (0.11)	0.56 (0.83)
	[0.09]	[0.15]	[0.03]	[0.18]	[0.10]
Age (β_3)	-0.12 (0.17)	0.13 (0.11)	0.10 (0.20)	-0.00 (0.05)	0.25 (0.41)
	[-0.05]	[0.07]	[0.03]	[-0.00]	[0.05]
Quality[a]	*p*=.006**	*p*=.012*	*p*=.007**	*p*=.261	*p*=.006**
Quality (= worse) (β_4)	-2.21 (0.59)	-1.09 (0.33)	-2.00 (0.60)	-0.01 (0.13)	-5.21 (1.46)
	[-0.87]**	[-0.61]**	[-0.65]**	[-0.01]	[-0.93]**
Quality (= univ.) (β_5)	1.56 (1.24)	0.93 (0.77)	2.50 (1.38)	0.57 (0.33)	4.85 (3.00)
	[0.62]	[0.52]	[0.81].	[0.69]	[0.87]
Condition[a]	*p* = .028*	*p* = .001***	*p* = .254	*p* = .000***	*p* = .000***
GAME (β_6)	0.88 (0.33)	0.85 (0.26)	0.59 (0.39)	0.91 (0.12)	3.36 (0.83)
	[0.35]*	[0.47]**	[0.19]	[1.10]***	[0.60]***
COMP (β_7)	0.14 (0.45)	0.17 (0.34)	0.27 (0.50)	-0.35 (0.16)	-0.11 (1.10)
	[0.06]	[0.09]	[0.09]	[-0.42]*	[-0.02]
$\hat{\sigma}_\eta$	0.791	0.291	0.773	0.000	1.958
$\hat{\sigma}_\xi$	0.000	0.427	0.001	0.223	1.079
$\hat{\sigma}$	2.527	1.766	3.082	0.795	5.522

Note. Standard errors are given in parentheses and effect sizes in brackets.

[a]Tests of significance of factor variables with three levels (likelihood ratio test).

†$p < .10$ *$p < .05$ **$p < .01$ ***$p < .001$

The results showed that different tests contributed differently to the between-condition differences in score1 and score2. The difference between the game media (EU-comp, EU-no-comp) and the non-game medium (Class) was most pronounced in the cases of epmem-score1 and epmem-score2 (large effect sizes), followed by negotiation-score1/2 (moderate effect sizes) and project-score2 (small to moderate effect size). No significant effect for the Class medium was found as concerns project-score1 and policy-score1/2. Nevertheless, in the case of all eight tests, participants' mean performance in the Class condition was worse or not better than participants' performance in the two game conditions. At the same time, in the case of no between-condition differences, the finding cannot be attributed to the floor effect, because mean performance of experimental participants, in any test, was significantly better than mean performance of naive participants (i.e., those not engaged in learning activities; see Sec. 4.2).

These outcomes can be interpreted as follows: the game media were most effective in promoting episodic memory and knowledge of the activity being practiced (i.e., negotiations). These are meaningful outcomes, yet we do not put much stock into these findings for two reasons. First, this part of the study was only exploratory. Second, the Episodic Memory Test was not a proper test of episodic/autobiographic memory. A proper test should have assessed the "what-where-when" aspects of events being experienced rather than just the "what" component. Such a test was beyond the scope of this work. Nevertheless, these findings may be useful for inspiring future studies; especially because better episodic memories can serve for cuing knowledge actually acquired. At the same time, the game conditions were not much more effective than the non-game condition in promoting learning of factual and/or conceptual knowledge. As concerns conceptual knowledge, this may seem, at first glance, to go against common intuition; especially because games and simulations are sometimes supposed to be particularly useful for teaching mental models (e.g., Brom et al., 2014). Nevertheless, mental model studies often focus on the acquisition of mechanical mental models rather than more hazy/ambiguous models for political concepts. In addition, our two game media in fact outperformed the non-game medium in the delayed Project Test assessing conceptual knowledge, and the effect sizes (small to moderate) were actually in the range of effect sizes reported by recent meta-analyses of media comparison game-based studies (Wouters et al., 2013; Sitzmann, 2011).

As concerns differences between the two game media, a significant effect for the EU-no-comp medium was found only in the case of epmem-score1/2: in favor of the EU-no-comp condition. This outcome can be explained by the fact that the EU-no-comp groups discussed, on average, fewer policies during a game run (Mean = 14.77, *SD* = 1.92), compared to both EU-comp (Mean = 16, *SD* = 0) and Class groups (Mean = 16.25, *SD* = 1.91). This is because there were fewer participants in the EU-no-comp groups (due to the procedure of assigning participants into the conditions). Thus, it could have been slightly easier for the EU-no-comp participants to remember and later recognize discussed agendas.

References

Brom, C., Bromová, E., Děchtěrenko, F., Buchtová, M., & Pergel, M. (2014). Personalized messages in a brewery educational simulation: Is the personalization principle less robust than previously thought? *Computers & Education, 72*, 339-366.

Sitzmann, T. (2011). A meta-analytic examination of the instructional effectiveness of computer-based simulation games. *Personnel Psychology, 64*(2), 489-528.

Wouters, P., van Nimwegen, C., van Oostendorp, H., & van der Spek, E. D. (2013). A Meta-Analysis of the Cognitive and Motivational Effects of Serious Games. *Journal of Educational Psychology, 105*(2), 249-265.