Notice

This is the author's version of a work that was accepted for publication in Computers & Education. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was accepted for publication in **Computers & Education (2017), DOI: 10.1016/j.compedu.2017.07.001**. The paper was accepted: **3-JULY-2017**.

Title page

1) Full title

Enjoyment or Involvement? Affective-Motivational Mediation during Learning from a Complex Computerized Simulation

2) Authors

Cyril Brom
Faculty of Mathematics and Physics, Charles University
Malostranské nám. 25, 118 00, Prague, the Czech Republic
brom@ksvi.mff.cuni.cz

Filip Děchtěrenko
Faculty of Mathematics and Physics, Charles University
Malostranské nám. 25, 118 00, Prague, the Czech Republic
filip.dechterenko@gmail.com

Nikola Frollová
Faculty of Mathematics and Physics, Charles University
Malostranské nám. 25, 118 00, Prague, the Czech Republic
nikola.frollova@gmail.com

Tereza Stárková

Faculty of Mathematics and Physics, Charles University

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

tereza.starek@gmail.com

Edita Bromová

Faculty of Mathematics and Physics, Charles University

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

edita@email.com

Sidney K. D’Mello

University of Notre Dame, Department of Psychology & Department of Computer Science

118 Haggar Hall, Notre Dame, IN 46556, United States

sdmello@nd.edu

3) Corresponding author

Cyril Brom

Faculty of Mathematics and Physics, Charles University, Room 312, Malostranske Namesti 25, Prague, 11800, Czech Republic.

E-mail: brom@ksvi.mff.cuni.cz

Tel: (420) 221 914 216; Fax: (420) 221 914 281

4) Source of funding

This study was primarily funded by Czech Grant Science Foundation (GA ČR) by Project nr. 15-14715S.

Sidney D’Mello was supported by the National Science Foundation (NSF) (DRL 1108845 and IIS 1523091). Any opinions, findings and conclusions, or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of NSF.

5) Acknowledgement

We thank Kateřina Svobodová from the Institute of Microbiology of the Czech Academy of Sciences, who helped us to prepare the citrate substrate simulation version. We thank research assistants who helped to conduct the experiments, most notably: Ondřej Smíšek, Kateřina Koppová, and Petra Šustová. We also thank Laboratory of Behavioral and Linguistic Studies in Prague, where part of the study was conducted.

Enjoyment or Involvement? Affective-Motivational Mediation during Learning from a Complex Computerized Simulation

Abstract

There is increased interest in augmenting multimedia instructional materials to elevate learners' positive affective-motivational states in order to improve learning. However, these efforts have only been partly successful and mediational effects of positive affective-motivational states have not always been established. In this study, university students (N = 65) from the Czech Republic, a country where beer brewing is a source of national pride, were informed that they would either study how to brew beer (high intrinsic motivation condition) or how to prepare a citrate substrate (low intrinsic motivation condition). The 90-minute simulation environment used for learning was about beer brewing in both cases, with superficial changes to instructions and graphics to disguise the topic manipulation. Generalized positive affect, overall enjoyment, flow, and learning involvement were higher in the beer brewing condition (Cohen's d = 0.44 – 0.87) as were learning gains when measured immediately (retention: d = 0.48; transfer: d = 0.46) and a month later (retention: d = 0.66; transfer: d = 0.62). However, only learning involvement and flow positively mediated the influence of the topic manipulation on immediate learning outcomes; there were no mediation effects on delayed learning outcomes after co-varying out immediate learning. The findings corroborate results from extant studies on the importance of topic interest in learning from instructional texts. They also indicate that affective-motivational mediation is one, but not the only, mechanism by which topic-based intrinsic motivation manipulations influence learning and that induced positive affective-motivational states can be differentially related or unrelated to learning.

Keywords: positive affect, flow, motivation, topic, multimedia learning

1. Introduction

Positive affective-motivational states of learners, such as situational interest, affective engagement, and enjoyment, play important roles in real-world learning contexts (e.g., Hidi & Renninger, 2006; Pekrun & Linnenbrink-Garcia, 2012; Lazowski & Hulleman, 2016). Thus, it stands to reason that learning can be enhanced by increasing the prevalence of these states (e.g., Keller, 2010; Lazowski & Hulleman, 2016; Yeager & Walton, 2011). One popular way to do this is by augmenting learning materials: a method called “emotional design” by some (Um et al., 2012; Plass & Kaplan, 2015; see also Norman, 2004).

The conditions under which “emotional design” actually facilitates learning have been most recently investigated in the context of digital game-based (e.g., Brom, Šisler, et al., 2016; Iten & Petko, 2016; Plass et al., 2013) and multimedia learning (e.g., Um et al., 2012; der Meij, 2013; Mayer & Estrella, 2014; Schneider, Nebel, & Rey, 2016). However, as detailed later, the findings are generally mixed. On the one hand, learners may be willing to invest more cognitive resources into learning when engaged by the augmented materials. On the other hand, the augmentations may serve as distractors or promote learning-irrelevant thinking.

There is a dearth of “emotional design” manipulations that elevate positive affective-motivational states while also improving learning. This is especially true for “minimalistic” manipulations that, beyond motivating and activating learners, only minimally alter how learners process materials. For instance, altering schematic graphics from an instructional animation such that certain graphical elements are anthropomorphized (while keeping their number, size, and positions intact; cf. Um et al., 2012) can be considered to be a “minimalistic” manipulation. In contrast, moving from an educational website to a quiz-based educational game with the same content (cf. Papastergiou, 2009) is not a “minimalistic” manipulation, because these two types of learning environments differ in numerous features (e.g., a game map, a simple plot, lives) that can influence cognitive processing.

From the theoretical and methodological perspectives, this complicates researching whether, and when, elevated positive affective-motivational states mediate the influence of “emotional design” manipulations on learning gains (i.e., affective-motivational mediation). From a practical perspective, educational designers are left without clear “emotional design” principles for designing appealing, yet instructionally effective, learning materials.

Our study addresses these challenges through two main goals. First, we develop a new, “minimalistic”, topic-based manipulation. This is important methodologically. Second, and more importantly, we use this manipulation to investigate affective-motivational mediation in the context of learning from a complex, 90-minute, computerized simulation. This is significant given that affective-motivational mediation during learning from multimedia instructional materials has been studied over much shorter learning sessions (e.g., Knörzer et al., 2015; Magner et al., 2014; Mayer & Estrella, 2014; Schneider et al., 2016; Um et al., 2012). Furthermore, we measure learning not only after the treatment, but also a month later. This enables us to investigate both immediate and delayed influences of positive affective-motivational states.

We experimentally manipulated the learning topic—such that one topic was expected to be more interesting and valuable compared to the other—to trigger intrinsic motivation. As far as we know, the influence of a motivating topic on learning has not yet been studied in the context of computerized simulations. However, because perceived task-value and relevance are known to play an important role in achievement conditions (see Eccles & Wigfield, 2002; Keller, 2010) and perceived topic interest affects learning from instructional texts (Fulmer, D'Mello, Strain & Graesser, 2015; Schiefele, 1999; Tobias, 1994), we expected our manipulation would trigger between-group differences in intrinsic motivation.

Our topic manipulation was as follows. In the Czech Republic, a country where beer brewing is a source of national pride, learners were informed that they would use a computerized simulation to either learn to brew beer (high intrinsic motivation condition) or to prepare a citrate substrate[1] (low intrinsic motivation condition). However, the simulation depicted beer brewing in both cases. For the citrate substrate version, only the title, certain key words in the instructions (e.g., yeast – fungal culture, acetone – toxin), and two graphical icons were replaced. The structure of the to-be-learnt process, the simulation's graphics and the level of graphical abstractness (with the two minor exceptions), the manner in which learners interacted with the simulation, the amount of extraneous information, and the linguistic style of instruction remained unaltered. In this sense, the manipulation can be considered to be "minimalistic" and should not be a source of distraction or contain seductive details (Mayer, 2009, ch. 4; Sweller et al., 2011; Garner et al., 1992).

Would we find that the topic influenced learning outcomes and would this relationship be affect-motivationally mediated? If so, by which variables? These are the central questions of the present study.

1.1. Cognitive-Affective Theory of Learning from Media and Cognitive Load Theory

The Cognitive-Affective Theory of Learning from Media (CATLM) (Moreno, 2005) is a common framework in multimedia learning and digital game-based learning literature. It is

[1] Citrate substrate is an existing substance. Citrates are chemical derivatives of citric acid. The term substrate refers to a bacterial growth medium. "Citrate substrate" is thus a bacterial growth medium consisting of certain derivatives of citric acid.

derived from the Cognitive Theory of Multimedia Learning (Mayer, 2009) and is related to Cognitive Load Theory (Sweller, Ayres, & Kalyuga, 2011; Kalyuga, 2011). CATLM assumes that a learner processes incoming information using his or her visual and verbal channels such that the learner first *selects* information from low-level sensory representations and then *organizes* the information in the form of visual and verbal mental models. In the next step, these models are *integrated* – together and also with prior knowledge from long-term memory – resulting in a cross-modal mental model represented in working memory. The crucial assumptions are that a) learners must actively engage in the selection, organization, and integration processes in order to construct a coherent mental model and b) learners have limited capacity in both channels and in working memory. With respect to the active engagement assumption (a), CATLM posits (apart from other things) that affective-motivational factors mediate learning by increasing or decreasing active cognitive engagement.

Cognitive Load Theory is related to the limited cognitive capacity assumption (b). Cognitive load (or working memory load) is the number of information elements that has to be simultaneously represented and processed in working memory. The theory, after its recent proposed adjustment by Kalyuga (2011), assumes two types of load during learning: *intrinsic* and *extraneous*.[2] Intrinsic load derives from the complexity of the learning task with respect to the learner's prior knowledge (what is complex for a novice may not be so complex for an expert). Learning would be suboptimal if intrinsic load were not managed by investing the appropriate amount of cognitive resources (e.g., because the learner is bored), or if the learner's cognitive capacity were overloaded. Extraneous load arises from suboptimal design

[2] Intrinsic and extraneous cognitive load are directly mapped onto the CATLM processes by Mayer (2009, pp. 79-89) and Kalyuga (2011, pp. 5-8).

of the learning environment, such that the learner must engage in processing that does not aid learning but is nevertheless invoked by the environment (e.g., if bits of texts are presented far apart or if extraneous information is present). Extraneous load thus hinders learning, or at least does not improve it (cf. Park, Flowerday, & Brünken, 2015), because it consumes cognitive resources that could otherwise be devoted to accommodating intrinsic load. If the learning environment contains elements that influence positively affective-motivational states, these elements can contribute positively to learning via affective-motivational mediation. However, they can also impair learning by increasing extraneous cognitive load.

Positive affective-motivational states can differentially influence cognitive processes beyond increasing active cognitive engagement. For instance, they can increase affect-congruent judgment effects, foster heuristic processing or alter the scope of attention (see Blanchette & Richards, 2010; Linnenbrink & Pintrich, 2004; see also Fiedler & Beier, 2014). Generally, it is assumed that these effects are positive with respect to learning, especially if the affective-motivational states are activating rather than deactivating (e.g. Pekrun & Linnenbrink-Garcia, 2012; see also Schiefele, 1999). However, they can negatively influence performance in certain contexts (Barth & Funke, 2010) or can even promote content-irrelevant thinking. Thus, they can also be considered sources of extraneous cognitive load (e.g., Knörzer, Brünken, & Park, 2016, Sec. 2.3.1; see also Pekrun & Linnenbrink-Garcia, 2012, p. 264).

Unfortunately, it is notoriously difficult to measure cognitive load; let alone distinguish between the two types of cognitive load (Brünken, Plass, & Leutner, 2003; Brünken, Seufert, Paas, & 2010; de Jong, 2010; see also Leppink et al., 2014). We address this by a) controlling for self-reported prior knowledge (because prior knowledge influences intrinsic cognitive load), including recruitment of low prior-knowledge learners only, b) measuring and controlling levels of overall cognitive load via a proxy variable: perceived difficulty, and c) attempting to equalize extraneous cognitive load across conditions (apart from influences due to affective-motivational activation).

1.2. The Affect-Motivational-Learning Link in Multimedia Learning and Digital Game-Based Learning Contexts

Certain socio-psychological motivational interventions in classrooms, such as writing assignments designed to enhance perceived relevance (Hulleman & Harackiewicz, 2009), improve learning interest and learning outcomes; especially for disadvantaged learners (see Yeager & Walton, 2011; Lazowski & Hulleman, 2016). However, it is less clear whether the same effects will occur from augmentations of multimedia learning materials designed to elevate positive affective-motivational states, particularly for learning materials that have already been “optimized” from the cognitive perspective using multimedia learning principles (Mayer, 2009; Clark & Mayer, 2011).

It has long been known that seductive details, which are interesting textual, pictorial or auditory additions that provide tangentially relevant information not necessary for comprehending the core instructional message, can harm, or at least, not improve learning. This is presumably because they induce extraneous cognitive load (see, e.g., Garner et al., 1992; Mayer, 2009; Rey, 2012). Therefore, researchers have searched for other ways to augment multimedia learning materials to evoke positive affective-motivational states (and facilitate learning).

As far as we know, the most promising findings in this regard suggest that the augmentations are beneficial for young children in the form of peer agent tutors in printed materials (der Meij, 2013) and in the form of certain game features; for instance, the presence of a narrative context in educational games (Cordova & Lepper, 1996; Sandberg, Maris, &

Hoogendoorn, 2014). For college students, these augmentations can also be beneficial in the form of anthropomorphizing certain graphical elements in short computerized multimedia learning materials (Um et al., 2012; Plass et al., 2014; Mayer & Estrella, 2014; Gong et al., 2017). It has also recently been shown that learning for this population was improved when decorative pictures (which are a form of seductive details) on large clickable buttons in multimedia learning environments were positively valenced or when they depicted learning rather than leisure contexts (Schneider, Nebel, & Rey, 2016).

Aside from these examples, the findings tend to be more mixed: with both null and negative effects documented in the literature (e.g., Adams et al., 2012; Brom et al., 2014; Brom, Hannemann, et al., 2016; van Dijk, 2010; Echevaría et al., 2012; Guo & Goh, 2016; Haaranen et al., 2015; Fassbender et al., 2012; Heidig et al., 2015; Schroeder & Taxler, 2017; Park et al., 2015; Vandercruysse et al., 2013). For instance, changing initial instruction (learning vs. entertainment) in a 3D VR simulation with an on-screen pedagogical agent can be beneficial for incidental conceptual learning, but only when paired with in-game quizzes with feedback (Erhel & Jamet, 2013). Also, the initial instruction manipulation may elevate positive affective-motivational states (Vadercruysse et al., 2013) or may fail to do so (Erhel & Jamet, 2013; see also Nebel et al., 2017). Affective embodied agents providing feedback in a learning game were shown to increase levels of affective-motivational states but not learning outcomes (Guo & Goh, 2016; see also Guo & Goh, 2015; Heidig & Clarebout, 2011). Likewise, an instructional video incorporating a human hand demonstrating how to solve physical problems may enhance ratings of the instructor’s engagement but hinder learning, compared to a video with no hand (Schroeder & Traxler, 2017).

The few studies that have investigated affective-motivational mediation report negligible-to-small correlations between affect and learning. One exception is a study with children who learnt from a printed tutorial enhanced with an agent, where induced flow, self-efficacy, and perception of task relevance all predicted learning gains (*r* = .26 - .52) (der Meij, 2013). Other findings are weaker. For instance, in Um and colleagues (2012), half the participants learnt from animated multimedia material with added visual anthropomorphisms[3]. Induced generalized positive affect, intrinsic motivation, and likability correlated mildly with test scores (*r* = .06 - .35). In an extension to this study (Plass et al., 2014), these correlations were even weaker, but they were higher for generalized positive affect and motivation (*r* = .17 - .28) compared to likability (*r* = .06 - .11).

In the context of learning from texts (i.e., without pictures), Schiefele (1999) reported an average correlation of .33 between ratings on the interestingness of narratives or text segments and various measures of text learning (based on a review of 14 text-learning studies). The relationship between interest and learning scores could be partly (but not fully) explained by more interested learners having higher prior knowledge (cf. Tobias, 1994; Schiefele, 1999, p. 271).

In the fields of multimedia and digital game-based learning, single-group correlational studies have tended to report negligible correlations between positive affective-motivational variables and learning outcomes (e.g., der Meij, 2014; Iten & Petko, 2014; Sabourin & Lester, 2014). There have also been differential relationships: for instance, Brom, Šisler, and colleagues (2016) reported that generalized positive affect, but not flow, mediated the effects of an educational game on learning outcomes. However, flow was predictive of learning outcomes in the study by der Meij (2013) mentioned above. Tangential to these findings, Schroeder, Romin, and Craig (2017) found that students' perceptions of a pedagogical agents' ability to be engaging was unrelated to learning outcomes.

[3] This study used a 2 x 2 design: present/absent anthropomorphisms vs. present/absent initial mood induction.

Overall, the findings suggest that the beneficial effects of a positive affective-motivational augmentation of multimedia learning materials on learning are variable and that the affect-learning relationship is somewhat fragile, though rarely negative. This also suggests that the field could benefit from more robust manipulations as the effects are weak. The field should also consider multiple positive affective-motivational states as there appear to be differential effects with respect to individual states.

1.3. Current Study

We experimentally manipulated the topic of a 90-minute interactive computerized simulation, such that one topic was designed to be more intrinsically motivating (beer brewing) than the other (citrate substrate production). We expected that the manipulation would trigger positive affective-motivational states that would positively influence learning. Learners can experience a number of affective-motivational states in academic contexts, such as they do when studying for an exam (e.g., Pekrun et al., 2009), or during technology-enhanced learning (e.g., Calvo & D'Mello, 2011; D'Mello, 2013) or when taking tests (e.g., Kleine et al. 2005) (see also Pekrun et al., 2002; Pekrun et al., 2011). Pekrun organizes these states along dimensions of valence (positive vs. negative), activation (activating vs. deactivating), and object focus (i.e., related to the undertaken activity or to the outcome, both prospectively and retrospectively) (Pekrun, 2006; Pekrun & Linnenbrink-Garcia, 2012). Some of these states can be experimentally induced prior to learning (e.g., Knörzer et al., 2016). Here, we focus on states induced during the learning session. We measure flow, learning involvement, generalized positive affect, and enjoyment, which can be conceptualized as activity-related, positive-activating states.

Flow is defined as pleasant absorption in an activity (Csikszentmihalyi, 1975), including increased attention to and concentration on the object of the activity (rather than on an achievement outcome). *Learning involvement* is also related to task concentration but involves positive feelings derived from learning. Whereas low *generalized positive affect* is related to deactivating positive states such as calmness or relaxation, high generalized positive affect is related to excitation or enthusiasm. *Enjoyment* can be viewed as an emotional state that occurs when the activity is positively valued and is sufficiently controllable by the learner (Pekrun, 2006; p. 323). Enjoyment derives from the activity undertaken per se, rather than its instrumental value.

The assignment to conditions was random, with gender and majors (computers science vs. social sciences and arts) balanced. We measured *personal value* and *task interest prior to learning* as manipulation checks, since these are related to intrinsic motivation (cf. Eccles & Wigfield, 2002; Hidi & Renninger, 2006; Keller, 2010; Ryan & Deci, 2000; Schiefele, 1999). To verify that the manipulation did not instigate differing expectancies for success, we measured *task anxiety prior to learning*. We also measured *perceived difficulty* during and after the learning session to ensure that the manipulation did not differentially influence task complexity (as perceived by learners). Because random assignment does not guarantee equivalence, we assessed whether the groups were balanced with respect to perceived prior knowledge, familiarity with key words in the instructions, prior generalized positive and negative affect, self-assessed knowledge of mathematics, frequency of playing videogames, frequency of playing live action experiential/simulation games, self-assessed ability of acquiring mental models, graphing skills and spatial visualization skills (see Figure 1).

We hypothesized that:

H1. Positive affective-motivational variables (i.e., flow, learning involvement, generalized positive affect, and enjoyment) would be higher in the beer brewing condition.

H2. Learning outcomes would be higher in the beer brewing condition.

H3. The positive affective-motivational variables will positively mediate the influence of the topic on learning outcomes.

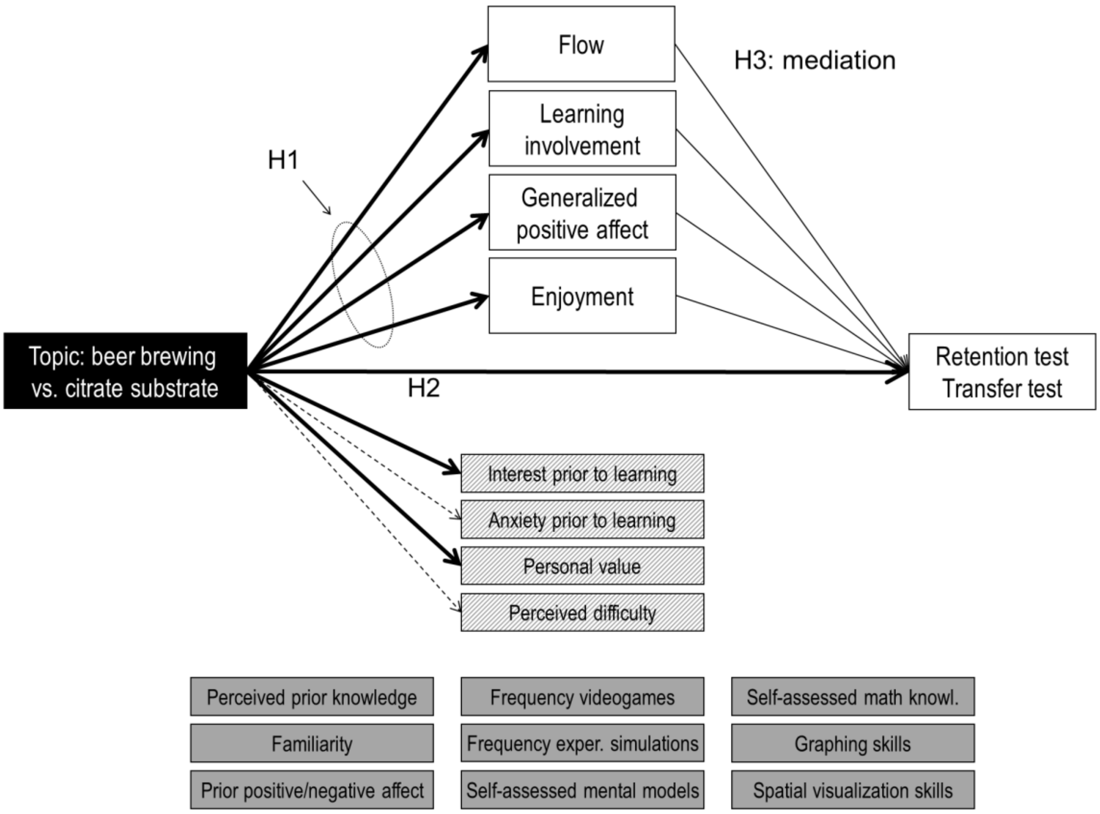

Figure 1. Manipulation (black), dependent (white), manipulation check (dashed), and control (gray) variables used in this study. Arrows represent key investigated causal influences. Bold arrows represent the expected positive influence of the beer brewing topic (with corresponding hypotheses highlighted). Thin arrows depict the influence of mediators. Dashed arrows correspond to no expected between-group differences.

2. Methods

2.1. Participants and Design

Seventy-four Czech university students with diverse backgrounds (in psychology, computer science, the arts, and new media studies) participated for course credit. The university was one of the highest ranked in the Czech Republic, suggesting that the participants were above average learners with respect to the general population. Generally, low prior knowledge of beer brewing could be expected in the sample (Brom et al., 2014). Low prior knowledge of citrate substrate production is also probable, because the topic is not covered by a typical high school or university with a non-chemistry curriculum. However, some learners could have higher prior knowledge of beer brewing (e.g., being home-brewers or fans) compared to citrate substrate production. To avoid possible bias, we excluded all high prior knowledge learners based on self-reported domain prior knowledge (as described in Sec. 2.4). In all, nine participants were excluded: one for high prior knowledge of beer brewing, one for being color-blind, three for technical issues, two for not comprehending the instructions, one for not feeling well, and one for negative prior attitude toward the experiment. Sixty-five participants (M_{age} = 23.60, SD_{age} = 3.75; 55% males) were included in the analysis.

The study used between-subject design with random assignment (with gender balanced) to a high (n = 30) vs. low (n = 35) intrinsic motivation group based on the simulation topic: beer (high) vs. citrate substrate (low). Computer science and social science (or art) students were evenly divided across conditions. Participants were unaware of the second condition. The groups were balanced with respect to age (p > .4; d = 0.2).

2.2. Materials – Simulation

We developed an interactive simulation on beer brewing using the Netlogo toolkit (Wilensky, 1999). The beer brewing process is rather complicated, and we focused only on the following four key phases which can be used by beginning home-brewers (see Brom et al., 2014 for details): infusion, hopback, fermentation and conditioning. The simulation was modified for the citrate substrate production condition as noted below. The simulation had three parts:

1. The *tutorial*, which demonstrated how to control the simulation (10-20 minutes).
2. The *linear part*, which demonstrated how to brew beer (30-50 minutes).
3. The *task-solving part*, in which the learner brewed several beers of a specific type in the simulation (30-40 minutes).[4]

The self-paced simulation was in the Czech language. Its graphical interface (Figure 2) included textual instructions, an animation panel showing the content of the preparation vessels used for brewing, an explanation panel relaying the meaning of graphical elements, panels showing the amount of ingredients in the product, an adjustable thermometer, buttons for controlling the processes, an "assessment" button providing learners feedback about how well they were doing, and a slider for controlling the simulation speed. Instructional texts were depicted on individual screens; the learner could return to previous screens.

The tutorial had 10 screens and the linear part had 24 screens. The task-solving part contained some of the instructions from the linear part (see below). The instructions had 3,467 words (3,410 for the citrate substrate version) (see Brom et al., 2014 for examples of

[4] The original version (detailed in Brom et al., 2014) has a fourth part not used here.

instructional texts). Several key ingredients were (visibly) present in the product, such as enzymes, starch or yeast (depending on the brewing phase). The amount of ingredients could be monitored through graphs and numerical panels. When the simulation was running, the graphs and panels were constantly updated, and the content of the vessel was animated. The simulation adhered to multimedia learning principles (Mayer, 2009) whenever possible to optimize the learning experience from a cognitive perspective.

Learners controlled each production phase with three buttons (start the phase, carry out the phase’s operation, end the phase). Each phase featured one operation which could be carried out repeatedly: adding malt in the first phase, adding hops in the second phase, adding yeast in the third phase, and adding sugar in the final phase. Learners could also adjust the temperature in the fermentation vessel at any moment, thereby changing the behavior of ingredients/elements in the vessel. It was possible to restart a phase if something went wrong.

Learners were guided during the tutorial and the linear part. They typically started by reading instructions from a particular screen to learn about the given portion of the process. The instructions assigned them a subtask on roughly half the screens; for example, boiling bacteria or adding hops in specific amounts at the appropriate time. Learners could inspect the consequences of their actions in the simulation environment and on graphs and numerical panels. They sometimes wrote down notes and/or requested feedback via the “assessment” button. Thereafter, learners proceeded to the next instruction screen.

Actions could be performed either when the simulation was running or when it was stopped. Sometimes, learners had to return to a previous instruction screen to understand the process better. Incorrect actions could sometimes be remedied. For instance, if a wrong temperature was set for a brief period during the fermentation phase, this could often (but not always) be corrected by adjusting the temperature. If a remedy was not possible, the learner could restart the phase or the whole process (for instance, when harmful bacteria started to produce acetone).

Subtasks were assigned in the tutorial and the linear part to provide learners with guidance. Learners could ignore these subtasks and could freely explore the simulation environment (which some of them did, often resulting in errors). During the task-solving part, instructions assigning subtasks were not displayed, but instructions pertaining to how the process worked were visible. Therefore, the element of guidance was eliminated in the task-solving part, increasing the possibility of choice and challenge. Learners could, however, still return to the instructional screens.

The citrate substrate version was also about beer brewing, but with certain superficial adjustments so that the process depicted citrate substrate production and the learners did not see through the manipulation. We replaced two figures (Figure 3), changed the initial instruction (Table 1), and replaced certain words or formulations from the instructions (Table 1) while also adjusting the sentences grammatically (e.g., changing the gender of a pronoun). All names of ingredients and biochemical products in the citrate substrate version (e.g. molasses, poisonous toxin) denoted existing elements, i.e., we did not use fictitious words. In a pilot with 35 participants, no overall difference in prior familiarity between the words from the replacement pairs was found, except for the final product terms (beer – citrate substrate; alcohol – citrate). As a partial remedy, we added a contextual explanation to the citrate substrate initial instruction so that the learners would imagine what the product is good for (see Table 1, the first line). We return to the issue of familiarity in Sections 2.3 and 4.

The groups were balanced with respect to time needed to complete the simulation ($M_{substrate}$ = 91.97, $SD_{substrate}$ = 17.28; M_{beer} = 92.85, SD_{beer} = 16.18; $p > 0.8$; $d = 0.05$).

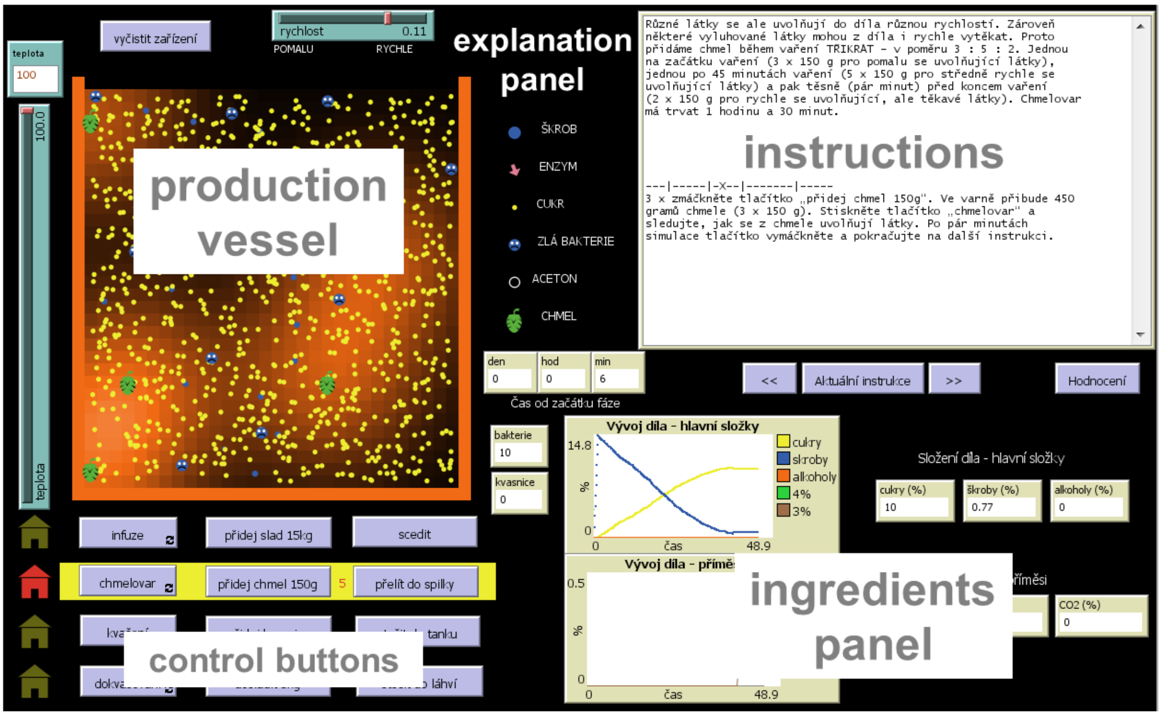

Figure 2. The simulation screenshot: the beginning of hopback. The temperature has been set to 100 degrees Celsius; the vessel is full of sugar (yellow dots) produced in the previous phase; hops (green cones) have been added and have started to release chemicals (orange “fog”); and harmful bacteria (blue spots with white “eyes”) are still present. The instructions are in the Czech language and they are depicted for illustrative purposes only (i.e., to demonstrate the interface’s layout).

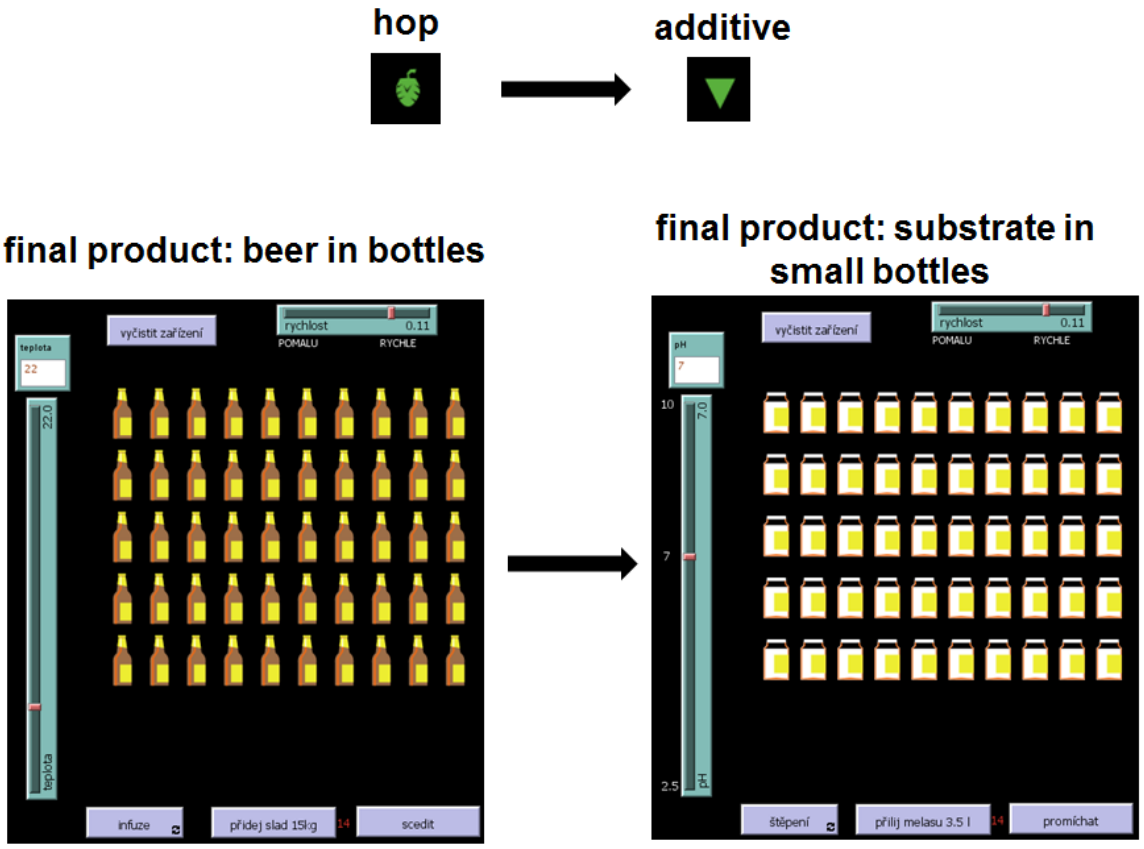

Figure 3. All graphical replacements (left: beer brewing; right: citrate substrate production).

Table 1

Replacements for the citrate substrate version

Group	Beer brewing	Citrate substrate
Initial instruction	"The goal of this simulation is to teach you how to brew **beer**, which is an **alcoholic** beverage produced by fermentation."	"The goal of this simulation is to teach you how to produce liquid **citrate substrate**, which you can imagine as **citric acid** mixed with sugar. It is used as a growth medium for microorganisms preferring acidic environments (i.e., **acidophilic microorganisms**). These microorganisms are then added to food (e.g., to **yogurts**) as **immune system boosters**. The microorganisms primarily feed on **citrate**, a key component of the citrate substrate."
Products	**beer**	**citrate substrate**
	ten-degree beer	substrate-ten
	twelve-degree beer	substrate-twelve
	stronger beer	more concentrated substrate
	alcohol	**citrate**
Ingredients, chemicals, intermediate products, by-products	**yeast**	**fungal culture**
	"harmful" bacteria	**"harmful" Penicillin**[a]
	hops	**additive**
	malt, i.e. germinated scrap barley	**molasses**, i.e., by-product of refining sugarcane into sugar
	acetone	**poisonous toxin**
	starch	**sucrose**
	hops in the form of cones, granules or hop extract	additive in the form of powder or granules
	fusel alcohol	**gloconate**
Production phases, processes	**infusion**	**breakdown**
	hopback	**enrichment**
	fermentation	**fermentation**[b]
	conditioning	**optimizing**
	brewing	producing
	surrogation	**nutrification**
	straining the product	mixing the product
	removing hopped wort	straining
	foam begins to turn brown	foam begins to turn black
	bottle the beer	bottle the citrate product
Quantities and values	**temperature**[c]	**pH**[d]
	8° C [optimal for yeasts]	pH 7 [optimal for Aspergillus]
	lower temperature [fine for	around neutral pH [fine for

	bacteria]	Aspergillus]
	0.5° - 30° C [livable for yeasts]	pH 3.5 – 9 [livable for Aspergillus]
	> 40° C [lethal for yeasts]	pH < 3 [lethal for Aspergillus]
	2° C [for conditioning phase]	pH 8 [for optimizing phase]
	75° C [for infusion phase]	pH 5 [for enrichment phase]
	100° C [for hopback phase]	pH 2.5 [for enrichment phase]
	1000 liters [volume produced]	100 liters [volume produced]
	5 kg sugar [for surrogation]	100 g sugar [for nutrification]
	20 min [until the bacteria dies in a 100° C product]	2 hrs. [until Penicillin dies in pH 2.5 product]
	90 min [the length of the hopback phase]	3 hrs. [the length of enrichment phase]
	several months or more [the length of the conditioning phase]	15 days or more [the length of optimizing phase]
	150 g of hop	100 g of additive
	1.5 % [malt per one-degree of beer]	3.5 % [molasses per one-degree of substrate]
Instruments	thermometer	pH regulator
	brewing tank[e]	tank
	lager tank	bioreactor
Taste-related	taste of beer	optimality/quality for the target microorganism
	tasty beer	beneficial for the metabolism of the target organism, i.e. “tasty”
	improving taste	balancing out of additives
	headache after drinking (and poor taste of) beer with fusel alcohol	“Ill” state of the target microorganism after consuming citrate substrate with gluconate (due to metabolic problems)
	taste beer [yourself]	taste citrate substrate [by a target microorganism]

Note: Words included in the familiarity test are in boldface. The words may have different degrees of familiarity in the Czech context and language compared to other cultural contexts. [a]University students are very familiar with Penicillin in the Czech Republic. [b]Two different words are used in the Czech language for “fermentation” for the two conditions. [c]The Celsius scale is used in the Czech Republic. Students should be familiar with the following values: 0° C = freezing point, 100° C = boiling point, around 20° C = room temperature. [d]Students should already be familiar with the pH scale (this topic is mandatory at high schools) and with the following values: pH 7 is neutral, lower pH levels (towards 0) are acidic, higher pH levels (towards 14) are basic; also, pH 5.5 is the standard skin pH (known from soap advertisements). [e]The Czech language uses word “varna”, which is less commonly-used in Czech than the term “brewing tank” is in English.

2.3. Procedure

Each session started between 9 am and 10 am. Participants were tested in groups of between one and four, each sitting at a separate computer in a lab. Participants in the same group received the same simulation version. The procedure is depicted in Figure 4. Participants were first informed that they would study a “particular process” with a computerized simulation (as a general introduction). They completed two questionnaires (familiarity and prior PANAS; see Section 2.4) and were then informed about the simulation topic (i.e., given the initial instructions from Table 1).

Participants were offered short breaks between the simulation parts (after filling in questionnaires associated with each part). They could take notes on sheets of paper handed to them, but we informed them that their notes would be taken away for the knowledge tests. They completed the immediate testing session after completing the simulation. After the testing session, they were interviewed. Roughly one month after the experiment, participants arrived for the delayed testing session (usually in the morning).[5]

[5] Participants were invited for “the second part of the experiment”, without being told that their knowledge would be tested. Many of them probably guessed it would. However, in a control question given after the delayed knowledge tests, they reported that they did not prepare for the second tests (M = 5.95, SD = 0.21; “To what extent did you prepare yourself for the tests?”; scale 1 – *very intensively*; 6 – *not at all*). The difference between groups was not significant (Mann-Whitney U Test: W = 529; p = .497).

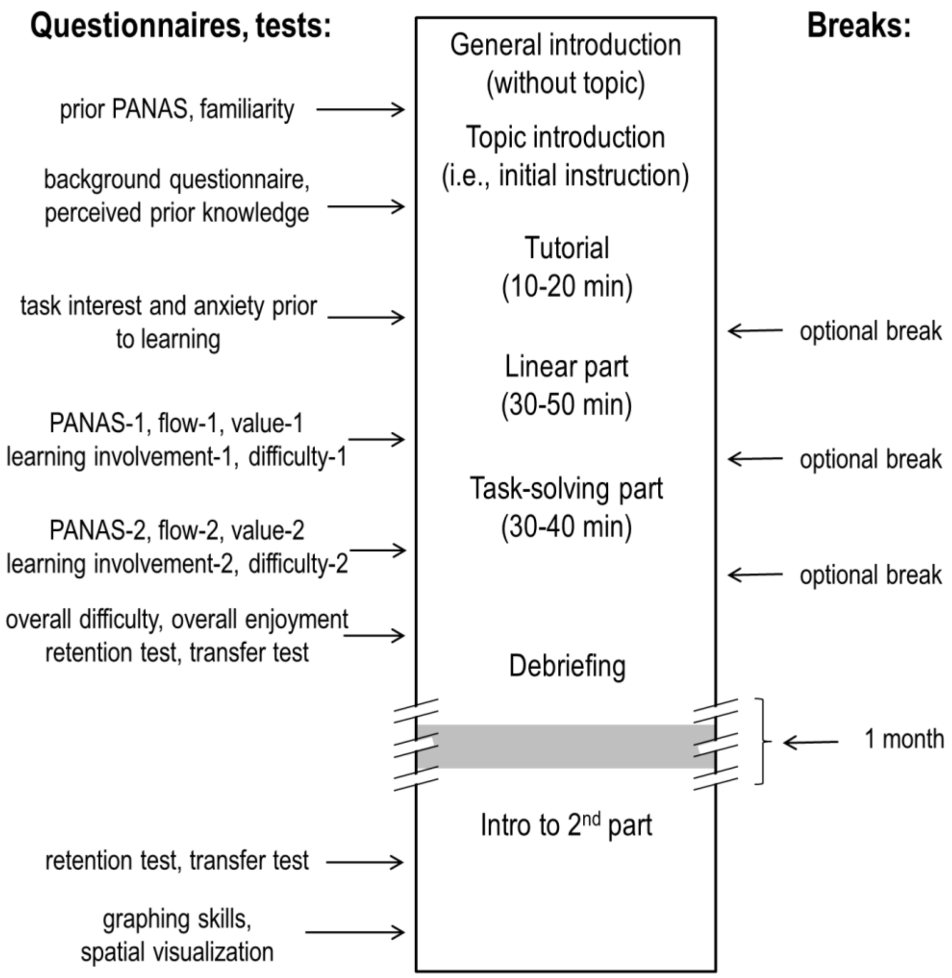

Figure 4. Experiment schedule.

2.4. Pen and Paper Materials

Learner **familiarity** with the key replaced words was assessed *before* they were informed about the topic. The questionnaire asked learners how familiar they were with the following words [a list of words] (5-point Likert scale). To contextualize the question, the learners were asked to imagine that they were preparing a chemical mixture in a high school laboratory and had encountered the words in the manual. For each topic, 13 topic-specific words were presented, along with 2 (beer brewing) or 6 (citrate substrate) key words from the initial instruction (see Table 1) and 32 (28) lures (beer brewing: $\alpha = .85$; citrate substrate: $\alpha = .90$).

The background questionnaire yielded information about the participant's age, gender, study type, native language, and vision difficulties. Based on Brom and Děchtěrenko (2015), it also included one question on self-assessed knowledge of mathematics (a 6-point Likert scale), one question on frequency of playing videogames (a 4-point ordinal scale), one on the frequency of playing live action experiential/simulation games (a 5-point ordinal scale), and one question on self-assessed ability to acquire mental models (a 7-point Likert scale). These questions were included to verify whether the groups were balanced (see Brom and Děchtěrenko, 2015 for details). For the same reason, we also administered (in a delayed testing session) a shortened version of a test on graphing in science (McKenzie & Padilla, 1986), according to Brom and colleagues (2014), and a 9-item version of a test on spatial visualization skills by Slezakova and Molnar (2011). The groups were balanced with respect to all of these variables (all ps > .4; all ds: $-0.2 < d < 0.2$).

Self-reported **domain prior knowledge** was measured (after the topic was introduced) in the beer brewing condition by a self-assessment pre-test based on Brom and colleagues (2014)**.**[6] The pre-test was created with an expert on beer-brewing and piloted on a sample of students. Participants used a checklist to report if any of the following applied: "My relatives (or I personally) brew beer", "I have taken part in an excursion to a brewery", "We learnt about beer brewing in school", "I know what *Saccharomyces cerevisiae* is", "I know how *Lactobacillus* can influence beer", "I know why malt is added to beer before yeast". The citrate condition used six complementary questions, e.g. "We learnt about bacterial fermentation in school." Participants received two points per checked item.

[6] Administering the full knowledge tests took around 30 minutes. Our pilots showed that administering full tests to low prior knowledge participants could be frustrating and demotivating. For this reason and also to avoid cueing participants on what they should remember, a knowledge test was avoided at this phase (cf., e.g., Mayer & Estrella, 2014, Um et al., 2012). Another issue was that the process of beer brewing is long and complex and cannot be assessed by a brief pre-test as is the case of simpler processes that can be learned in a couple of minutes and are often used in the multimedia learning literature (e.g., Park, Flowerday, & Brünken, 2015).

Participants were also instructed to write down whether they had ever attempted to learn about beer brewing/citrate substrate production. This open-ended question was scored by one rater and was awarded 0 – 4 points based on a predefined key (the answers were repetitive, and it was thus possible to construct such a key based on our previous work (Brom et al., 2014); e.g. “I saw one TV document.” – 2 points; “I saw two TV documents.” – 3 points).

Participants were then asked to assess their knowledge of why and how alcohol is created during the beer brewing process (on a 4-point ordinal scale, i.e., a multiple-choice question; the score was scaled down to 0 – 2 to match the checklist questions above). The complementary question asked them to assess their knowledge of why and how acids are produced during growth of fungi.

Finally, participants answered two questions (on a 6-point Likert scale) pertaining to the following: their self-reported knowledge of beer brewing or citrate substrate production (0 – 5 points) and whether they could explain why a morning headache can be worse when drinking non-alcoholic beer rather than alcoholic beer the evening before (for the beer brewing condition) or why spoiled homemade fruit wine has a sour, vinegary taste (for citrate substrate production condition) (the score was scaled down to 0 – 2 points). Thus, the total range on the pre-test was 0 – 25. Cronbach's α was .71[7]. Participants with 12.5 or more points were excluded on the basis of having high prior knowledge.

[7] The total pre-test score comprises four questions with ordinal or Likert scales and six yes/no questions. Each of these yes/no questions yields either 0 or 2 points. To compute α, scores from these six yes/no questions were tallied and the sum was used as one variable in the calculation of α (along with the other four variables).

Task interest prior to learning was assessed with five questions from the Questionnaire on Current Motivation[8] (Rheinberg, Vollmeyer, & Burns, 2001). The questions were as follows: “Today’s topic seems to be very interesting to me”, “I am eager to see how I will perform on the today’s task”, “I’m really going to try as hard as I can on this task”, “While doing this task, I will enjoy discovering how to prepare [beer | citrate substrate]”, “I would work on this task even in my free time (if I have this instructional simulation)” (7-point Likert scale) (α = .85).

Task anxiety prior to learning was assessed with three questions from the same questionnaire: “When I think about the task, I feel somewhat concerned”, “I’m afraid I will make a fool of myself”, “I think I won’t do well at the task” (7-point Likert scale) (α = .82). We were interested in task interest and task anxiety after the participants became familiar with both the topic and the simulation, so we measured these two variables after participants completed the tutorial (see Figure 4).

We administered the PANAS (Positive and Negative Affect Schedule; Watson, Clark, & Tellegen, 1988) to measure participants’ **generalized affective state**. It consists of two, 10-item mood scales: one for positive and the other for negative affect (5-point Likert scale). The PANAS was administered before the topic introduction and twice during the intervention (positive affect: α = .85, .88, .87; negative affect: α = .74, .80, .93). The instruction from the prior PANAS addressed current feelings; whereas, in the in situ measurements they referred to feelings from the previous simulation part (i.e., “mark to what extent you experienced these feelings during **the previous part** of the simulation experience” [emphasis in the original]).

[8] The original instrument was too long for our purposes, and we consistently used these eight questions in our research. The questions load on two factors: interest and anxiety. Confirmatory factor analysis showed a good fit for our present data with these two factors (χ^2/df = 1.39; CFI = .97; TLI = .95; RMSEA = .078).

Flow was measured twice (α =.85, .90) with ten 7-point Likert items from the Flow Short Scale (Rheinberg, Vollmeyer, & Engeser, 2003) (e.g., “I do not notice time passing”, “I feel I have everything under control”). The instruction was related to the previous part of the simulation.

Along with this questionnaire, we also administered one **personal value** question (“Today’s activity is personally important to me”) with the same scale for use as a manipulation check. Because we were interested in whether personal value is stable during learning, this question was also administered twice. The two values correlated highly (r = .87) and the scores were averaged for the subsequent analysis.

Learning involvement was assessed twice (α = .85, .83) with a researcher-created questionnaire consisting of eight 7-point Likert items that were inspired by various questionnaires on motivation/interest/involvement-related constructs (e.g., Schraw, Bruning, & Svoboda, 1995; Isen & Reeve, 2005): “So far, I’m enjoying [topic]”, “I was always sure what I was supposed to do next”, “I always knew how to complete the assigned tasks”, “I’m tired”, “I’m looking forward to the next part” (the 2nd administration: “I’d like to continue in the [topic’s] activity”), “I focused on the [topic’s] activity”, “I think I am doing well so far”, “I was careful and conscientious when completing the tasks.”. The instruction was also related to the previous simulation part.

Participants also completed the following **perceived difficulty** question along with this questionnaire: “The difficulty of the simulation [2nd administration: task-solving] meets my expectations”; 7-point Likert scale: *much more difficult – much less difficult*.

Participants received a feedback questionnaire after the learning session, but before the knowledge tests were administered. Four items from this questionnaire were relevant to the present research. Three of the items pertained to **overall enjoyment**: “I enjoyed doing this activity", "I thought this was a boring activity", "I would describe this activity as very interesting" (6-point Likert scale; α = .74). These items were taken from the interest/enjoyment subscale of the Intrinsic Motivation Inventory (McAuley, Duncan, & Tammen, 1987). This construct complemented generalized positive affect, flow, and learning involvement in that it related to the overall learning experience. The fourth question was on **overall perceived difficulty**: "How difficult was today's lesson on [topic] for you?" (6-point Likert scale). The score from this question was averaged with scores from the two difficulty questions administered during the experiment (after converting them to the 1-6 scale) (α = .67).[9]

The knowledge test**s** were researcher-created and based on Brom and colleagues (2014) with minor adjustments. Each test had two complementary versions (one for immediate assessment and the other for delayed assessment; the versions were counterbalanced across participants). The tests were the same for both topics except for the replacement of certain key words pertaining to the topic.

The knowledge tests were composed of a retention and a transfer test. The **retention** test had 11 questions: one open-ended, one multiple-choice, and nine short answer questions. Examples include the following: "Write down the names of the four main phases of [topic] production in the correct order, as you learnt them today.", or "In what phase or phases of [topic] production are enzymes present during the whole phase?". The questions were scored by one rater based on a predefined exact key (e.g., the correct name of the phase – 1 point; a partially correct name [based on a pre-specified list of all names accepted as partially correct] – 0.5 point). The exact key was also constructed for the open-ended question (the presence of

[9] In these questionnaires, we also piloted 17 questions irrelevant to the present study; these are not discussed further.

certain keywords in the answer was awarded 1 point or 0.5 point). The score range was 0 – 32.

One version of the **transfer** test had 6, and the other 7, open-ended questions, e.g.: “How does fermentation differ from [beer: conditioning / citrate substrate: optimization]? Describe all the differences that occur to you.”, or “Why does the chance that the product will spoil increase, if we cannot manage a stable [beer: temperature / citrate substrate: pH] over the whole fermentation phase? Explain in detail.”. Unlike for the retention test, it was not possible to create an answer key that would be exact. The rating system was thus based on detection of key idea units in the answers (cf. Mayer, 2009). For each question, a list of idea units was available from a previous study (Brom et al., 2014). Participants were awarded 1 point for each correct idea unit or 0.25 or 0.5 points for a partially-correct idea unit.

The first author scored all the answers. To check for confirmation bias, a second independent rater (who was hypothesis-blind) also scored the answers. Correlations between raters’ scores for individual questions ranged from .91 to 1.00, which reflects substantial agreement. We also computed the inter-rater agreement with weighted Cohen’s κ (with weights set to zero on the diagonal and to the squared distance off the diagonal, where weights indicated the magnitude of the raters’ disagreement; Cohen, 1986). Cohen’s κ were in the .96 - .98 range for total scores and .79 - 1.0 for individual questions, which reflects substantial agreement. Raters’ scores were averaged for the subsequent analysis. The score range was 0 – 18 or 19 based on the version.

2.5. Analysis

Because we used two versions of the transfer tests, we z-transformed scores from each version to obtain comparable values. In two cases, one transfer test question was not given to a participant due to a technical problem. We replaced the missing value with the average score for the participant (for the given subtest). Flow data was converted to T-norms provided with the standardized Flow Short Scale (Rheinberg, 2004) (final scale 21-74). Flow, generalized positive affect, and learning involvement scores from the two measurements were strongly correlated (*r* = .70, .85, .75) and were thus averaged for the subsequent analyses.

Data was analyzed with the R statistical software suite (R Core Team, 2016). Primary between-group comparisons were conducted using independent-samples t-tests (two-sided) and effect size, measured using Cohen's *d*, was classified as small (*d* ~ 0.2), medium (*d* ~ 0.5) or large (*d* ~ 0.8) (Cohen, 1988). When a covariate was included, we used a one-way ANCOVA with effect sizes expressed using η_p^2 and classified as small ($\eta_p^2 = .01$), medium ($\eta_p^2 = .06$) or large ($\eta_p^2 = .13$) (Cohen, 1988). Confidence intervals for the effect sizes of the ANCOVA were bootstrapped (*N* = 1,000). For clarity, we also report converted η_p^2 values to Cohen's *d* using the formula from Cohen (1988) and an R package *compute.es* (AC Del Re, 2013). Pearson's correlation coefficients were classified into small (*r* ~ .1), medium (*r* ~ .3) and large (*r* ~ .5) (Cohen, 1988). Tests of assumptions for the statistical tests are described in the respective subsections.

3 Results

3.1. Manipulation checks

We note that personal value was significantly and substantially higher for the beer brewing compared to the citrate substrate group, suggesting that the topic manipulation had its intended effects (see Table 2). Task interest measured after the tutorial was also higher in the beer brewing group, though the effect was only modest and marginally significant (*p* =

.061). Task interest was measured after the tutorial, i.e., when participants had already interacted with the simulation for about 15 minutes, which might thus explain marginal effect. Several participants indicated in the debriefing that even though they did not highly value the citrate substrate production task, learning from the simulation was still modestly interesting for them (primarily because of its interactivity). Finally, anxiety and perceived difficulty did not differ between groups; suggesting no confounding influences of the manipulation.

Table 2. Manipulation check variables: means, standard deviations, and t-test.

Variable[a]	Condition		t-test				
	Citrate substrate	Beer brewing	*t*	*p*-value	*df*	*d*	95 % CI *d*
personal value [1 – 7]	2.64 (1.29)	4.20 (1.48)	4.52	.000[d]	63	1.13	[0.59, 1.66]
task interest [5 – 35]	22.8 (5.41)	25.4 (5.42)	1.91	.061	63	0.47	[-0.03, 0.98]
task anxiety [3 – 21][b]	9.14 (4.11)	9.23 (4.61)	0.08	.934[d]	63	0.02	[-0.48, 0.52]
perceived difficulty [1 – 6][c]	3.28 (0.76)	3.28 (0.70)	0.00	.997	63	0.00	[-0.50, 0.50]

[a]Possible ranges are given in square brackets.

[b]Higher values mean “higher anxiety”.

[c]Higher values mean “more difficult”.

[d]Personal value and task anxiety were not normally distributed (Shapiro-Wilk’s test; *p*s $< .016$). However, the results of the non-parametric Mann-Whitney test were similar (personal value: $p < .001$; task anxiety: $p = .984$).

3.2. Prior Knowledge and Prior Generalized Affect

Table 3 shows that the groups were balanced with respect to prior positive and negative affect (i.e., before the topic was introduced). However, self-reported domain prior knowledge was modestly higher in the beer brewing condition. Because this variable was somewhat correlated with some of the dependent variables (in particular, with immediate transfer: $r = .20$, $p = .06$; and with delayed transfer: $r = .18$, $p = .08$), we included it as a covariate in the subsequent analyses. Familiarity with key words was also marginally higher in the beer brewing condition ($p = .091$). However, this variable did not correlate with any of the dependent variables, so it was not analyzed further.

Table 3. Initial knowledge and prior generalized affect: means, standard deviations, and t-test.

Variable[a]	Condition		t-test				
	Citrate substrate	Beer brewing	*t*	*p*-value	*df*	*d*	95 % CI *d*
prior positive affect [10 – 50]	25.8 (6.92)	25.9 (5.48)	0.11	.915	62	0.03	[-0.47, 0.52]
prior negative affect [10 – 50]	13.2 (3.17)	13.6 (3.61)	0.43	.668[d]	62	0.11	[-0.39, 0.60]
self-reported domain prior knowledge [0 – 25]	1.93 (1.97)	3.44 (3.30)	2.27	.027[d]	63	0.56	[0.06, 1.07]
familiarity [0 – 4][b, c]	2.87 (0.50)	3.08 (0.41)	1.72	.091[d]	52[c]	0.43	[-0.07, 0.93]

[a]Possible ranges are given in square brackets.

[b]Higher values mean “higher familiarity”.

[c]Due to technical issues, familiarity scores for 11 participants are missing (4 for the beer brewing and 7 for the citrate substrate condition).

[d]Prior negative affect, self-reported domain prior knowledge, and familiarity were not normally distributed (Shapiro-Wilk’s test; *p*s < .001). However, the results of the non-parametric Mann-Whitney test were similar (prior negative affect: $p = .886$; self-reported domain prior knowledge: $p = .055$, familiarity: $p = .080$).

3.3. Hypothesis 1: The Effect of Topic on Affective-Motivational Variables

Table 4 shows that induced flow, learning involvement, and overall enjoyment were significantly higher in the beer brewing condition with medium to large effect sizes. Generalized positive affect was marginally ($p = .081$) higher in the expected direction and there was no difference for generalized negative affect. Because flow, learning involvement, generalized positive affect, and overall enjoyment were correlated (Table 5), we analysed this data with a MANCOVA[10]. This yielded a significant main effect for condition (Pillai test; $F(1, 63) = 4.22$; $p = .005$), thereby supporting Hypothesis 1.

Table 4. Affective motivational variables: means, standard deviations, and ANCOVA with self-reported domain prior knowledge as a covariate. Negative affect is shown for the sake of completeness.

Variable[a]	Condition		ANCOVA[d]					
	Citrate substrate	Beer brewing	F	p-value	df	η_p^2	95 % CI η_p^2	d[c]
positive affect [10 – 50]	28.7 (7.21)	32.1 (6.48)	3.15	.081	1, 61	.05	[.00, .18]	0.44
negative affect [10 – 50]	13.5 (3.91)	12.9 (3.46)	0.15	.698[d]	1, 61	.00	[.00, .03]	-0.10
flow [21 – 74][b]	53.4 (6.83)	57.3 (7.86)	4.83	.032	1, 61	.07	[.00, .22]	0.55
learning involvement [8 – 56]	42.1 (6.88)	46.9 (5.32)	8.04	.006[d]	1, 61	.12	[.02, .25]	0.70
enjoyment [1 – 6]	4.83 (0.69)	5.40 (0.50)	12.4	.001	1, 61	.17	[.04, .33]	0.87

[a]Possible ranges are given in square brackets.

[b]Scale after the transformation through T-norms.

[c]Corresponding Cohen's d for the η_p^2.

[d]Bartlett's test showed equal variances in both groups for all variables (ps > .078). Negative affect and learning involvement showed deviation from normality (Shapiro-Wilk's test; ps < .010). ANCOVAs are robust to the normality violation (Levy, 1980; see also Schmider et al., 2012), but the results can be biased if normality is violated due to presence of outliers (see Stevens, 2012). Using the Grubb's test, we detected one outlier for negative affect ($p = .030$) and one for learning involvement ($p = .007$). Also, self-reported domain prior knowledge, used as a covariate, had one outlier ($p = .019$). We therefore excluded these outliers and re-ran the tests—the results were very similar (positive affect: $p = .075$, negative affect: $p = .926$; flow: $p = .031$; learning involvement: $p = .009$; enjoyment: $p = .001$).

[10] Assumptions of the MANCOVA were met, except for the multivariate assumption of normality, which was narrowly missed ($p = .048$). However, MANCOVAs are robust with respect to deviances from the normal distribution (Olson, 1974).

Table 5. Correlation matrix for dependent variables. Number in parentheses denotes number of values used for correlation.

	positive affect	negative affect	flow	learning involvement	enjoyment	retention test immediate	retention test delayed	transfer test immediate
positive affect	-							
negative affect	-0.16 (65)	-						
flow	0.55*** (65)	-0.48*** (65)	-					
learning involvement	0.61*** (65)	-0.49*** (65)	0.84*** (65)	-				
enjoyment	0.60*** (65)	-0.27* (65)	0.55*** (65)	0.60*** (65)	-			
retention test immediate	-0.06 (65)	-0.07 (65)	0.31* (65)	0.37** (65)	-0.09 (65)	-		
retention test delayed	0.05 (63)	-0.24† (63)	0.31* (63)	0.40** (63)	-0.00 (63)	0.65*** (63)	-	
transfer test immediate	0.01 (65)	-0.20 (65)	0.33** (65)	0.42*** (65)	-0.06 (65)	0.68*** (65)	0.68*** (63)	-
transfer test delayed	0.05 (64)	-0.11 (64)	0.30* (64)	0.38** (64)	-0.04 (64)	0.65*** (64)	0.77*** (63)	0.66*** (64)

† *p* < .1 * *p* < .05 ** *p* < .01 *** *p* < .001

3.4. Hypothesis 2: The Effect of Topic on Learning Gains

Table 6 shows that all scores from the knowledge tests were consistently higher for the beer brewing condition, with somewhat larger differences for the delayed rather than immediate knowledge assessments. Because of inter-correlations (Table 5), we statistically analysed the difference using MANCOVA with self-reported prior knowledge as a covariate[11]. The main effect for the topic was marginally significant for the immediate tests ($F(1, 61) = 2.59$; $p = .083$) and was significant for delayed tests ($F(1, 59) = 4.7$; $p = .013$).

[11]Assumptions of the MANCOVA were met, except for the multivariate assumption of normality in the case of delayed tests ($p = .001$). However, MANCOVAs are robust with respect to deviances from the normal distribution (Olson, 1974).

We can thus conclude that Hypothesis 2 was supported; with marginal effects for immediate test scores potentially attributable to the sample size.

Effects of the topic on delayed learning outcomes after co-varying out immediate learning were marginally significant, and they were of the same magnitude as the effects of the topic on immediate learning outcomes (Table 6; Lines 3 and 6). This suggests that the topic influenced not only initial knowledge acquisition, but also how knowledge was retained between the two measurement points.

Table 6. Learning outcome variables: means, standard deviations, and ANCOVAs with given covariates.

Variable[a]	Covariate	Condition		ANCOVA[c]					
		Citrate substrate	Beer brewing	*F*	*p*-value	*df*	η_p^2	95 % CI η_p^2	*d*[b]
retention immediate [0 – 31]	prior knowledge	22.8 (5.05)	24.9 (3.99)	3.77	.057[c]	1, 61	.06	[.00, .18]	0.48
retention delayed [0 – 31]	prior knowledge	13.4 (6.08)	18.0 (6.71)	7.07	.010	1, 59	.11	[.00, .30]	0.66
retention delayed	prior knowledge, retention immediate			3.79	.057	1, 55	.06	[.00, .25]	0.48
transfer immediate [Z-scores]	prior knowledge	0.17 (0.93)	0.67 (0.84)	3.52	.066	1, 61	.05	[.00, .20]	0.46
transfer delayed [Z-scores]	prior knowledge	-0.70 (0.86)	-0.08 (0.87)	6.40	.014	1, 60	.10	[.00, .26]	0.62
transfer delayed	prior knowledge, transfer immediate			3.64	.062	1, 56	.06	[.00, .22]	0.47

[a]Possible ranges are given in square brackets.

[b]Corresponding Cohen's d for the η_p^2

[c]Test assumptions were met, except for the normality assumption in the case of immediate retention (Shapiro-Wilk's test; *p* = .010). ANCOVAs are robust to the normality violation (Levy, 1980), but the results can be biased if normality is violated due to the presence of outliers (see Stevens, 2012). No outlier was detected (Grubb's test: *p* = 0.17). Because self-reported domain prior knowledge, used as the covariate, had one outlier (*p* = .019), we re-ran the tests without this outlier. The results were very similar (retention immediate/delayed/delayed with immediate as a covariate: *p* = .058, .011, .066; transfer immediate/delayed/delayed with immediate as a covariate: *p* = .068, .014, .068).

3.5. Hypothesis 3: Affective-Motivational Mediation

We tested whether the influence of the topic on a) initial learning and b) retention of knowledge between two measurement points was mediated by any of the positive affective-motivational variables (i.e., induced flow, generalized positive affect, learning involvement, and overall enjoyment). Concerning point (a), all analyses included self-reported domain prior knowledge as a covariate. Concerning point (b), analyses also included scores on the respective immediate test as an additional covariate. Mediator models (i.e., the effect of topic and covariates on a positive affective-motivational variable) and outcome variable models (i.e., the effect of the topic, covariates, and a mediator on a learning outcome variable) were expressed as linear regressions. The mediation analysis was conducted using the package *mediation* (Tingley, Yamamoto, Hirose, Keele, & Imai, 2014) for causal mediation. As recommended by Preacher and Hayes (2004), we computed estimates for the indirect effect using quasi-Bayesian Monte Carlo simulations (N = 10,000). Although new procedures for estimating mediations do not require having a significant direct effect (for discussion, see: Zhao, Lynch Jr. & Chen, 2010), in our case all direct effects were at least marginally significant. Table 7 shows the results of the mediation analysis.

The main findings were that learning involvement and flow mediated the influence of the topic on immediate learning outcomes. Overall enjoyment marginally mediated the influence of the topic on immediate retention test scores, but the direction of mediation was reversed, suggesting that this effect might be spurious. There were no mediation effects for the delayed tests (after co-varying out both initial learning and self-reported domain prior knowledge).[12] This indicates that the experimental manipulation influenced how knowledge was retained between the two measurement points not only by elevating affective-motivational states indexed by our instruments, but also by other means (discussed in Section 4.2).

We can thus conclude that Hypothesis 3 was only partially supported: with respect to learning involvement/flow and immediate knowledge tests.

[12] When the mediation model included only self-reported domain prior knowledge, but not the immediate test scores, i.e., when the effects of immediate learning were not accounted for, flow mediated the effects of the topic on delayed test scores marginally (retention: Mean estimate = 0.13; $p = .090$; transfer: Mean estimate = 0.13; $p = .058$) and learning involvement significantly (retention: Mean estimate = 0.23; $p = .012$; transfer: Mean estimate = 0.20; $p = .030$).

Table 7. Mediation of the effect of the topic on test scores by positive affective-motivational variables.

Variable	Test	Covariate	Mean estimate[a]	95 % CI	*p*-value[c]
positive affect	retention immediate	prior knowledge	-0.05	[-0.21, 0.07]	.404
	retention delayed	prior knowledge, retention immediate	0.02	[-0.07, 0.14]	.626
	transfer immediate	prior knowledge	-0.03	[-0.19, 0.09]	.616
	transfer delayed	prior knowledge, transfer immediate	0.01	[-0.09, 0.13]	.858
flow	retention immediate	prior knowledge	0.15	[0.00, 0.36]	.048
	retention delayed	prior knowledge, retention immediate	0.03	[-0.08, 0.16]	.502
	transfer immediate	prior knowledge,	0.16	[0.00, 0.37]	.048
	transfer delayed	prior knowledge, transfer immediate	0.03	[-0.06, 0.14]	.604
learning involvement	retention immediate	prior knowledge	0.23	[0.04, 0.49]	.008[b]
	retention delayed	prior knowledge, retention immediate	0.07	[-0.05, 0.22]	.278[b]
	transfer immediate	prior knowledge,	0.25	[0.05, 0.51]	.002[b]
	transfer delayed	prior knowledge, transfer immediate	0.04	[-0.08, 0.18]	.526[b]
enjoyment	retention immediate	prior knowledge	-0.20	[-0.50, 0.02]	.074
	retention delayed	prior knowledge, retention immediate	-0.04	[-0.25, 0.15]	.682
	transfer immediate	prior knowledge	-0.17	[-0.44, 0.03]	.112
	transfer delayed	prior knowledge, transfer immediate	-0.09	[-0.34, 0.11]	.350

[a]Coefficients are mean estimates of indirect path.

[b]Learning involvement and self-reported domain prior knowledge had one outlier each. The results were similar when we re-ran the tests without these outliers (retention immediate/transfer immediate: p = .004, .004; retention delayed/transfer delayed: p = .278, .308).

[c]Self-reported domain prior knowledge had one outlier. The results were similar when the tests were re-run without this outlier (positive affect: p > .490; enjoyment: p > .110; flow ← retention delayed/transfer delayed: p > .416; flow ← retention immediate: p = .060, Mean estimate = 0.16, 95 % CI: [-0.00, 0.39]; flow ← transfer immediate: p = .048, Mean estimate = 0.16, 95 % CI: [0.00, 0.41]).

4. Discussion

4.1. Main Findings

Our primary aim was to investigate affective-motivational mediation using a topic-based manipulation in the context of studying a complex process from a 90-minute-long computerized simulation. By presenting learners with a more or less intrinsically motivating topic (beer brewing vs. citrate substrate production), but teaching them the same underlying process, we sought to influence how induced generalized positive affect, flow, learning involvement, and overall enjoyment influenced performance on immediate and delayed, retention and transfer tests.

We found that all four positive affective-motivational variables were consistently higher for learners studying the more motivating beer-brewing topic, though some effects were only modest (Table 4). The data also suggests that even learners who studied the less motivating topic were quite engaged, presumably because learning from a complex, interactive simulation was novel and interesting to those in our sample (i.e., above average university students). In fact, several participants, especially (but not only) those with a computer science background, noted this in the interview. The positive effect of interfacing with the simulation was enhanced when paired with the more motivating beer brewing topic.

We also found that both retention and transfer test scores were higher for students studying the more motivating topic. This effect was observed immediately after the learning sessions as well as a month later (Table 6). The effect sizes were modest for immediate tests and in the medium-to-large range for delayed tests. Performance on the delayed learning assessment tended to be superior with medium effect sizes for students studying the more motivating topic even after co-varying out immediate test scores. This suggests that the manipulation influenced not only the quality of initial learning, but also long-term knowledge retention. Our findings are consistent with research on learning from texts, which suggests that that interestingness of text fragments/topics predicts learning gains, especially at deeper levels of comprehension (see Schiefele, 1999; Tobias, 1994). However, of the four positive affective-motivational variables, only learning involvement and flow positively mediated the influence of the topic on learning gains; and only on the immediate tests. This means that in addition to being differentially related to learning, the positive affective-motivational variables we measured did not account for some of the between-group difference in learning outcomes; especially as concerns long-term retention of knowledge.

This study also had a methodological goal of developing a new “minimalistic” topic-based manipulation that beyond motivating and activating learners, minimally altered how learners processed the materials. This was done by changing the topic and a few words or formulations in the instructions, while keeping the structure of the to-be-learnt process, language style of instructional texts, how learners interacted with the simulation, the amount of extraneous information, and the simulation’s graphics intact. In our opinion, this study achieved this methodological goal, though several suggestions for improvements are discussed in the next sections.

4.2. Interpreting the Results

From the perspective of theories linking positive affective-motivational factors to improved cognitive processing, it is not surprising that higher involvement and flow (rather than generalized positive affect or enjoyment) mediated immediate learning gains. Of the constructs measured, learning involvement and flow are most closely linked to increased cognitive activation posited to be the underlying causes of improved learning (e.g., Moreno, 2005). It is plausible that learners can be focused, involved in learning, and perceive success (as indexed by learning involvement and partly by flow) but have neutral rather than positive feelings (as indexed by enjoyment and generalized positive affect). Such situations would presumably lead to better learning outcomes compared to situations in which learners are not fully involved in learning but enjoy the experience and have positive feelings towards it. Indeed, liking is not the same as learning.

Alternatively, more enjoyment and positive feelings could have facilitated learning-relevant cognitive processes beyond learning involvement and flow, but these additional improvements could have been countered by negative influences. For instance, a learner in the beer brewing condition could be emotionally excited (as indexed by generalized positive affect/overall enjoyment), which might increase cognitive activation but also engender learning-irrelevant thinking, e.g., prospectively thinking about discussing new knowledge about beer brewing with friends. It might also be the case that learners with better self-regulation abilities (not measured here) could achieve better learning outcomes even when the learning process is not particularly enjoyable and/or associated with positive feelings.

In the field of game-based learning, it is assumed that higher activation/motivation induced by playing serious games should be more beneficial in the long-term (cf. Brom et al., 2011; Brom, Šisler, et al., 2016). However, we found no mediation effects on long-term retention of knowledge (i.e., on the delayed tests after co-varying out initial learning). It is possible that the influence of the topic on knowledge retention might be attributed to cognitive rather than affective-motivational factors. First, higher personal relevance of the beer brewing topic could trigger different (and better) encoding processes than the less relevant topic. For instance, beer brewing can be better connected to the self-structure, a powerful processing and encoding cognitive-affective entity (cf. Symons & Johnson, 1997), via numerous autobiographical experiences with drinking beer or seeing others drinking beer, and therefore information can be better retained due to the mnemonic abilities of the self-structure. Second, the beer brewing topic could trigger more thinking/discussions about the topic in the month between assessments, leading to rehearsal effects. Third, there may have been between-group differences in prior schemata that were not captured by our prior knowledge and perceived difficulty instruments (cf. Tobias, 1994). These schemas, when activated by the more motivating beer brewing topic, would differentially influence intrinsic cognitive load, which could lead to better long-term information encoding for the beer brewing group. However, we cannot exclude the possibility that the influence of the topic on knowledge retention might be attributed to affective-motivational factors. Either learners' self-report ratings of generalized positive affect, flow, learning involvement, and enjoyment might not be representative of their actual states, or there could be an affective-motivational state accounting for the knowledge retention effect that we failed to measure in the study.

To what extent do our results generalize to different contexts? Even though our findings are consistent with recent theories on the affect-learning link, some caution is needed here. The weak relationship between enjoyment or generalized positive affect and learning gains is common in multimedia and game-based learning (Iten & Petko, 2014; Plass et al., 2013; see also Barth & Funke, 2010). However, positive correlations between these variables and learning outcomes, though small to medium only, have also been reported (e.g., Plass et al., 2014; Schneider et al., 2016). In short, the findings reported in the literature seem to be unstable. This might be because the measurement error that accompanies self-reported data masks true effects or positive affective-motivational constructs are differentially related to learning gains in different contexts. These alternatives will need to be addressed in future research, perhaps through a meta-analysis of studies that utilize diverse measures of positive affective-motivational states beyond self-reports and in a range of contexts.

4.3. Limitations and Suggestions for Future Studies

This study is not without limitations. First, even though our sample size is not atypical for an experimental multimedia learning study, some of our between-group differences were only marginally significant. A larger sample would increase statistical power. In our case, we do not think marginally significant findings pose much of a problem, because all the between-group differences trended in the same direction. Nevertheless, given small-to-medium effect sizes have often been found in multimedia and game-based learning "emotional design" studies, larger samples may be needed for this type of research.

Second, our study did not reveal why there was no affective-motivational mediation in the case of long-term knowledge retention; our explanations are speculative. Given delayed learning outcomes are rarely measured in "emotional design" research in general, we believe this point highlights the importance of investigating delayed learning gains (and their causes) in future studies on multimedia learning.

Third, our "minimalistic" experimental manipulation depends on the quality of the disguise in the citrate substrate condition. During the debriefing, several participants noted that the process of fake citrate substrate production reminded them of an alcohol production process though none said that the process masked beer brewing. Furthermore, even though we controlled for self-reported prior knowledge, word familiarity, and perceived difficulty, we cannot exclude the possibility that the manipulation indirectly influenced participants' intrinsic cognitive load, because of the difficulty in measuring this construct (see Section 1). Therefore, it would be useful for a future study to investigate whether our pattern of results can be replicated with another "minimalistic", topic-based manipulation in the context of multimedia learning. It would be also useful if unobtrusive intrinsic load measurements were developed in the future. The same applies for developing strategies for assessing prior knowledge of complex/lengthy processes in a manner that is not demotivating or frustrating for low prior knowledge learners.

Fourth, we measured learners' familiarity with words in word-replacement pairs but not familiarity with the products of beer vs. citrate substrate *per se*. Thus, we do not know whether the results were caused because of higher familiarity with beer as opposed to citrate substrate, or because of the higher value of/interest in the former product compared to the latter; or because of both. Future research using our type of "minimalistic" manipulation might consider including a control topic of a supposedly less valuable/interesting, but equally or even more familiar, product (e.g., making dish-washing detergent). This could help untangle the effects of interestingness/value from familiarity of the product. Furthermore, we used an authentic beer brewing process as the basis for deriving the second condition. However, an artificial process can be used for both conditions. This has methodological consequences as concerns development of this type of "minimalistic" manipulation: on the one hand, this could simplify designing the simulations and make them even more similar; on the other hand, ethical considerations come into play as we could be teaching inaccurate information.

Fifth, this study focused only on a selection of positive-activating, activity-oriented affective motivational states (cf. Pekrun & Linnnenbrink-Garcia, 2012). Other affective-motivational states, both positive and negative, can also influence learning. For instance, D'Mello and colleagues (2014) demonstrated that confusion can enhance learning under the appropriate circumstances (i.e., when appropriately regulated and resolved). Also, positive (Um et al., 2012) as well as negative (Knörzer et al., 2016) mood induction procedures can have positive impact on subsequent learning under specific conditions.

Finally, self-reports on affective-states can be complemented by different measurement techniques, such as automated facial emotion recognition procedures (Bosch et al., 2016) or retrospective affective judgement protocols, among other methods to measure affect during learning (Porayska-Pomsta et al., 2013).

4.4. Closing Remarks

This study demonstrated consistent effects of topic interestingness/value on positive affective-motivational outcomes as well as on learning gains; extending the results of text-based interest studies (Schiefele & Krapp, 1996; Schiefele, 1999; see also Fulmer et al., 2015) to a 90-minute, complex, multimedia simulation. From a practical perspective, this study re-emphasizes the importance of increasing perceived value: either by appealing to utility (e.g., Hulleman & Harackiewicz, 2009), interest and relevance (e.g., Keller, 2010) or through some other mechanism. The methodological contribution of this work is the development of a new, “minimalistic”, topic-based manipulation. On a more theoretical level, the study suggests that affective-motivational mediation may only be one mechanism by which intrinsic value manipulations influence learning. The hunt is on for additional mechanisms, especially when it comes to long-term retention of knowledge.

References

AC Del Re (2013). compute.es: Compute Effect Sizes. R package version 0.2-2, Available: http://cran.r-project.org/web/packages/compute.es (Accessed 16 December 2016).

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.

Barth, C. M., & Funke, J. (2010). Negative affective environments improve complex solving performance. *Cognition and Emotion, 24*(7), 1259-1268.

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.

Bosch, N., D'Mello, S., Baker, R., Ocumpaugh, J., & Shute, V. (2016). Using video to automatically detect learner affect in computer-enabled classrooms. *ACM Transactions on Interactive Intelligent Systems, 6*(2), 17.11-17.31.

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.

Brom, C., & Děchtěrenko, F. (2015). Mathematical Self-Efficacy as a Determinant of Successful Learning of Mental Models From Computerized Materials *ECGBL-9th European Conference on Games Based Learning* (pp. 89-97): Academic Conferences and publishing limited.

Brom, C., Hannemann, T., Stárková, T., Bromova, E., & Dechterenko, F. (2016). Anthropomorfic faces and funny graphics in an instructional animation may improve superficial rather than deep learning: A quasi-experimental study. *Proceedings of European Conference on e-Learning* (pp. 87 - 97).

Brom, C., Preuss, M., & Klement, D. (2011). Are educational computer micro-games engaging and effective for knowledge acquisition at high-schools? A quasi-experimental study. *Computers & Education, 57*(3), 1971-1988. doi:10.1016/j.compedu.2011.04.007

Brom, C., Šisler, V., Slussareff, M., Selmbacherová, T., & Hlávka, Z. (2016). You like it, you learn it: affectivity and learning in competitive social role play gaming. *International Journal of Computer-Supported Collaborative Learning, 11*(3), 313-348.

Brunken, R., Plass, J. L., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 53-61.

Brünken, R., Seufert, T., & Paas, F. (2010). Measuring cognitive load *Cognitive load theory* (pp. 181-202).

Calvo, R. A., & D’Mello, S. K. (Eds.) (2011). *New perspectives on affect and learning technologies* (Vol. 3). Springer Science & Business Media.

Clark, R. C., & Mayer, R. E. (2011). *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning* (3rd ed.): John Wiley & Sons.

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.

D'Mello, S. (2013). A selective meta-analysis on the relative incidence of discrete affective states during learning with technology. *Journal of Educational Psychology, 105*(4), 1082-1099.

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

Donald, A. N. (2004). *Emotional design: Why we love (or hate) everyday things*: Basic Books.

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual review of psychology, 53*(1), 109-132.

Echeverría, A., Barrios, E., Nussbaum, M., Améstica, M., & Leclerc, S. (2012). The atomic intrinsic integration approach: A structured methodology for the design of games for the conceptual understanding of physics. *Computers & Education, 59*(2), 806-816.

Erhel, S., & Jamet, E. (2013). Digital game-based learning: Impact of instructions and feedback on motivation and learning effectiveness. *Computers & Education, 67*, 156-167.

Fassbender, E., Richards, D., Bilgin, A., Thompson, W. F., & Heiden, W. (2012). VirSchool: The effect of background music and immersive display systems on memory for facts learned in an educational virtual environment. *Computers & Education, 58*(1), 490-500.

Fiedler, K., & Beier, S. (2014). Affect and cognitive processes in educational contexts. . In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 36-56). New York, NY: Routledge.

Fulmer, S. M., D'Mello, S. K., Strain, A., & Graesser, A. C. (2015). Interest-based text preference moderates the effect of text difficulty on engagement and learning. *Contemporary educational psychology, 41*, 98-110.

Garner, R., Brown, R., Sanders, S., & Menke, D. J. (1992). "Seductive Details" and Learning from Text *The role of interest in learning and development* (pp. 239-254).

Gong, S., Shangguan, C., Zhai, K., Guo, Y. (2017). The effects of emotional design on multimedia learning. *Acta Psychologica Sinica*, *49*(6), 771-782.

Guo, Y. R., & Goh, D. H. L. (2015). Affect in Embodied Pedagogical Agents: Meta-Analytic Review. *Journal of Educational Computing Research, 53*(1), 124-149.

Guo, Y. R., & Goh, D. H. L. (2016). Evaluation of affective embodied agents in an information literacy game. *Computers & Education, 103*, 59-75.

Haaranen, L., Ihantola, P., Sorva, J., & Vihavainen, A. (2015). In search of the emotional design effect in programming *IEEE/ACM 37th IEEE International Conference on Software Engineering* (Vol. 2, pp. 428-434): IEEE.

Heidig, S., & Clarebout, G. (2011). Do pedagogical agents make a difference to student motivation and learning? *Educational Research Review, 6*(1), 27-54.

Heidig, S., Müller, J., & Reichelt, M. (2015). Emotional design in multimedia learning: Differentiation on relevant design features and their effects on emotions and learning. *Computers in Human behavior, 44*, 81-95.

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

Hulleman, C. S., & Harackiewicz, J. M. (2009). Promoting interest and performance in high school science classes. *Science, 326*(5958), 1410-1412.

Isen, A. M., & Reeve, J. (2005). The influence of positive affect on intrinsic and extrinsic motivation: Facilitating enjoyment of play, responsible work behavior, and self-control. *Motivation and emotion, 29*(4), 295-323.

Iten, N., & Petko, D. (2014). Learning with serious games: is fun playing the game a predictor of learning success? *British Journal of Educational Technology, 47*(1), 151-163.

Kalyuga, S. (2011). Cognitive load theory: How many types of load does it really need? *Educational psychology review, 23*(1), 1-19.

Keller, J. M. (2010). *Motivational design for learning and performance: The ARCS model approach*: New York: Springer.

Kleine, M., Goetz, T., Pekrun, R., & Hall, N. (2005). The structure of students' emotions experienced during a mathematical achievement test. *ZDM (Zentralblatt füur Didaktik der Mathematik), 37*(3), 221-225.

Knörzer, L., Brünken, R., & Park, B. (2016). Facilitators or suppressors: Effects of experimentally induced emotions on multimedia learning. *Learning and Instruction, 44*, 97-107.

LaBar, K. S., & Phelps, E. A. (1998). Arousal-mediated memory consolidation: Role of the medial temporal lobe in humans. *Psychological Science, 9*(6), 490-493.

Lazowski, R. A., & Hulleman, C. S. (2016). Motivation Interventions in Education A Meta-Analytic Review. *Review of Educational Research, 86*(2), 602-640.

Leppink, J., Paas, F., Van Gog, T., van Der Vleuten, C. P., & Van Merrienboer, J. J. (2014). Effects of pairs of problems and examples on task performance and different types of cognitive load. *Learning and Instruction, 30*, 32-42.

Levy, K. J. (1980). A Monte Carlo study of analysis of covariance under violations of the assumptions of normality and equal regression slopes. *Educational and Psychological Measurement*, *40*(4), 835-840.

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). London: Lawrence Erlbaum.

Magner, U. I., Schwonke, R., Aleven, V., Popescu, O., & Renkl, A. (2014). Triggering situational interest by decorative illustrations both fosters and hinders learning in computer-based learning environments. *Learning and Instruction, 29*, 141-152.

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

Mayer, R. E., & Estrella, G. (2014). Benefits of emotional design in multimedia instruction. *Learning and Instruction, 33*, 12-18.

McAuley, E., Duncan, T., & Tammen, V. V. (1989). Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: A confirmatory factor analysis. *Research quarterly for exercise and sport, 60*(1), 48-58.

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

Nebel, S., Schneider, S., Schledjewski, J., & Rey, G. D. (2017). Goal-Setting in Educational Video Games: Comparing Goal-Setting Theory and the Goal-Free Effect. *Simulation & Gaming, 48*(1), 81–97.

Norman, A. D. (2004). *Emotional design: Why we love (or hate) everyday things*: Basic Books.

Olson, C. L. (1974). Comparative robustness of six tests in multivariate analysis of variance. *Journal of the American Statistical Association, 69*(348), 894-908.

Park, B., Flowerday, T., & Brünken, R. (2015). Cognitive and affective effects of seductive details in multimedia learning. *Computers in Human Behavior, 44*, 267-278.

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., Elliot, A. J., & Maier, M. A. (2009). Achievement goals and achievement emotions: Testing a model of their joint relations with academic performance. *Journal of Educational Psychology, 101*(1), 115.

Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary educational psychology, 36*(1), 36-48.

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational psychologist*, *37*(2), 91-105.

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

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., & Kaplan, U. (2015). Emotional design in digital media for learning *Emotions, technology, design, and learning* (pp. 131-162): Academic Press.

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

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior research methods, instruments, & computers, 36*(4), 717-731.

Porayska-Pomsta, K., Mavrikis, M., D'Mello, S. K., Conati, C., & Baker, R. (2013). Knowledge elicitation methods for affect modelling in education. *International Journal of Artificial Intelligence in Education, 22*, 107-140.

R Core Team (2016). A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/ (Accessed 15 October 2016).

Rheinberg, F. (2004). *Motivationsdiagnostik [Motivation diagnosis]*: Gottingen: Hogrefe

Rheinberg, F., Vollmeyer, R., & Burns, B. D. (2001). FAM: Ein Fragebogen zur Erfassung aktueller Motivation in Lern-und Leistungssituationen [QCM: A questionnaire to assess current motivation in learning situations]. *Diagnostica*, *47*, 57-66.

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.

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

Sandberg, J., Maris, M., & Hoogendoorn, P. (2014). The added value of a gaming context and intelligent adaptation for a mobile learning application for vocabulary learning. *Computers & Education*, *76*, 119-130.

Schiefele, U. (1999). Interest and learning from text. *Scientific studies of reading*, *3*(3), 257-279.

Schiefele, U., & Krapp, A. (1996). Topic interest and free recall of expository text. *Learning and individual differences, 8*(2), 141-160.

Schmider, E., Ziegler, M., Danay, E., Beyer, L., & Bühner, M. (2010). Is it really robust? Reinvestigating the Robustness of ANOVA Against Violations of the Normal Distribution Assumption. *Methodology, 6*, 147-151.

Schneider, S., Nebel, S., & Rey, G. D. (2016). Decorative pictures and emotional design in multimedia learning. *Learning and Instruction, 44,* 65-73.

Schraw, G., Bruning, R., & Svoboda, C. (1995). Sources of situational interest. *Journal of Literacy Research, 27*(1), 1-17.

Schroeder, N. L., Romine, W. L., & Craig, S. D. (2017). Measuring pedagogical agent persona and the influence of agent persona on learning. *Computers & Education*, *109*, 176-186.

Schroeder, N. L., & Traxler, A. L. (2017). Humanizing Instructional Videos in Physics: When Less Is More. *Journal of Science Education and Technology*, *26*(3), 269-278.

Slezakova, J., & Molnar, J. (2011). Testování geometrické představivosti [“Testing geometrical imagery”], *E-Pedagogikum, 3,* 80-96. [in Czech]

Stevens, J. P. (2012). *Applied multivariate statistics for the social sciences*. Routledge.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York: Springer.

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). Mediation: R package for causal mediation analysis. *Journal of Statistical Software*, *59*(5).

Tobias, S. (1994). Interest, prior knowledge, and learning. *Review of Educational Research, 64*(1), 37-54.

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 der Meij, H. (2014). Developing and testing a video tutorial for software training. *Technical Communication, 61*(2), 110-122.

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.

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.

Wilensky, U. (1999) NetLogo, Center for Connected Learning and Computer-Based Modeling, Northwestern University, available: http://ccl.northwestern.edu/netlogo/ (Accessed 16 October 2016).

Yeager, D. S., & Walton, G. M. (2011). Social-psychological interventions in education They're not magic. *Review of Educational Research, 81*(2), 267-301.