Anthropomorphic Faces and Funny Graphics in an Instructional Animation May Improve Superficial Rather than Deep Learning: a Quasi-Experimental Study

Cyril Brom[1], Tereza Hannemann[2], Tereza Stárková[1,2], Edita Bromová[1,3] and Filip Děchtěrenko[1]
[1]Faculty of Mathematics and Physics, Charles University in Prague, Czech Republic
[2]Faculty of Arts, Charles University in Prague, Czech Republic
[3]Film and TV School, Academy of Performing Arts in Prague, Czech Republic
brom@ksvi.mff.cuni.cz
tersel@seznam.cz

Abstract: Information about what visual elements in instructional animations enhance learning via mediating effects of elevated engagement is largely lacking. In this study, high school students (n = 41) interacted in a laboratory with a roughly 6-minute-long, black-and-white, instructional animation with "emotionally" enhanced graphics. The topic was biological wastewater treatment. The enhancements included a) adding static faces for two schematic visual elements and b) changing the neutral appearance of a fish and a river bed to funny appearances. The participants' learning outcomes (assessed by retention and transfer tests) and the participants' state engagement (indexed by generalized positive affect and flow levels) were compared to data from a group of comparable students interacting with the same animation with neutral graphics (n = 37) in a previous study. The two groups did not differ in state engagement (flow: d = -0.16; positive affect: d = 0.03) and transfer test scores (d = 0.13), but there was a small trend favoring the enhanced graphics in the retention test, when corrected for pretest scores (η^2 = 0.038; p = .095). Qualitative data suggest that the graphical enhancements might serve as memory cues during the test phase for some participants. The small effect of the enhanced graphics on retention may thus be of cognitive, rather than of affective, origin. This study demonstrates the importance of considering emotional manipulations in future research.

Keywords: Multimedia learning; Animations; Emotional Design; Flow; Positive Affect; E-learning.

1. Introduction

Instructional animations are an important member of the family of educational tools and a notable example of e-learning technology. They can be used in various ways; for instance, by students who self-study at home or by teachers showing them to the class on a projector. From a theoretical perspective, animations are computerized multimedia learning materials: they combine text and pictures (Mayer, 2009) and use computers as a presentation medium.

Principles exist for how to design multimedia learning materials, including animations, so that their instructional message can be cognitively processed by learners in an effective way (Mayer, 2009; Clark & Mayer, 2011). These principles are not only empirically-based, they are also theoretically well organized by the frameworks of the Cognitive Theory of Multimedia Learning (Mayer, 2009) and Cognitive Load Theory (Sweller, Ayres, & Kalyuga, 2011). No complementary principles are available for how to design multimedia learning materials so that they are not only cognitively optimal, but also affectively appealing; further enhancing learning (i.e., emotional design principles; Plass & Kaplan, 2016). Absence of such principles is a notable drawback because students' affective states are quite important in educational contexts and they influence learning (e.g., Pekrun, 2006), decision making (e.g., Isen, 2001), and memory (e.g., Linnenbrink & Pintrich, 2004). In fact, only a handful of studies have addressed the issue of emotional design in the context of multimedia learning (see, e.g., Mayer, 2014; Park et al., 2015; Heidig et al., 2015). Even the field of educational computer games, which can be viewed as multimedia learning materials in a broader sense, has limited knowledge of what game features create an affectively optimal experience that enhances learning (see Wouters et al., 2013).

One possible emotional design principle is "to anthropomorphize graphical elements in the materials" (cf. Park et al., 2015; Mayer & Estrella, 2014). An example of this principle is adding human-like faces and/or expressive eyes to otherwise non-human entities. Recent research has demonstrated some benefits of this design principle both in affective and cognitive terms; especially when a gray-scale or a black-and-white visualization is enhanced by warm, bright colors (Um et al., 2012; Plass et al., 2014, Exp. 1; Mayer & Estrella, 2014; but see also Plass et al., 2014, Exp. 2). Another possible emotional design principle is "to change entities' graphical appearance from a neutral one to a funny one". This design principle is intuitively used by the designers of many entertainment games. However, whether or not it is beneficial for learning remains unclear (Snetsinger & Grabowski, 1993).

The present study investigates whether a subtle pictorial manipulation in a 6-minute-long, black-and-white animation increases learners' state engagement and thereby enhances learning. We emotionally enhanced the neutral visual appearance of an already existing instructional animation on the functioning of a wastewater treatment plant in the following ways: A) we added static human-like faces to two, otherwise schematically portrayed, chemical elements; B) we changed the neutral appearance of a mutated fish to a funny appearance; and C) we changed the neutral appearance of garbage at the river bed to a funny one. We used no colors and the added faces were static (i.e., without expressive eyes or mouth).

The study thus compares learning from two versions of the same animation: one version with a neutral visual appearance and the other with an emotionally enhanced visual appearance. Specifically, for this study, we recruited participants for the enhanced condition and compared their results to those of participants who interacted with the animation's neutral version in a different study (Brom et al., submitted). Would the enhanced graphics be beneficial for participants' engagement and learning results?

2. Theoretical Background

Traditionally, the design principles of multimedia learning (e.g., Mayer, 2009; Clark & Mayer, 2011) focus on the cognitive aspects of processing learning messages rather than the materials' affective appeal (cf. Mayer, 2014; Um et al., 2012; Heidig et al., 2015). Likewise, the related theories, such as the Cognitive Theory of Multimedia Learning (Mayer, 2009), focus primarily on the cognitive dimension of learning. Yet over a dozen affective states that students experience during learning by means of new technologies have been robustly identified (reviewed in D'Mello, 2013). Only recently, affective dimensions have started to be considered by multimedia learning researchers.

State engagement, also called engaged concentration (e.g., Baker et al., 2010), is an important affective state. As far as we know, it has so far not been operationalized precisely, but it has been tentatively linked to mild generalized positive affect and certain components of flow state: such as focused and intense attention.[1] As such, it can be indexed by a Positive Affect Schedule (Watson, Clark, & Tellegen, 1988) and a flow inventory such as Flow Short Scale (Rheinberg, Vollmeyer, & Engeser, 2003). We use these two constructs in this study, i.e., generalized positive affect and flow levels, as proxies to state engagement.

As a theoretical basis for linking state engagement to learning effects, we use an "affective" expansion of the Cognitive Theory of Multimedia Learning: the Cognitive-Affective Theory of Learning with Media (CATLM) (Moreno, 2005). In a simplified way, the CATLM posits that the effectiveness of learning depends on the effectiveness of selecting relevant information by the learner from the instructional message, its organization into a coherent mental model in the learner's working memory and the integration of this model into the learner's prior knowledge base. The efficiency of these processes depends, to a certain extent, on the amount of cognitive resources allocated for them. The amount of cognitive resources, in turn, depends on some other variables; including the level of actual engagement in the learning process (i.e., the state engagement). The higher state engagement is, the more cognitive resources will be allocated for the activity being undertaken: up to a certain point.

However, there is the following trade-off: manipulation of multimedia learning material can increase state engagement and thereby the amount of allocated cognitive resources (enhancing learning). However, at the same time, it can also require a substantial amount of cognitive resources for the cognitive processing of this manipulation (hindering learning). A prime example of an engaging manipulation that hinders learning under many circumstances is seductive images (i.e., illustrative images adding engaging information that is only weakly relevant for the core learning message) (Mayer, 2009). Therefore, multimedia learning researchers have started looking at how to alter learning materials to induce and/or to keep high state engagement during learning, while minimally distracting the learner from the cognitive processing of the instructional message's core. This effort can be called an emotional design research program (cf. Plass & Kaplan, 2015).

[1] The relationship between positive affect and flow state, on the one hand, and state engagement (engaged concentration), on the other hand, was pointed out to us by Sidney D'Mello [email correspondence from 9 March 2014].

The emotional design research program is in its infancy. As concerns changing neutral visualizations to funny ones, we are aware, as of now, only of a study by Snetsinger and Grabowski (1993), which reported null results. As concerns adding anthropomorphisms in combination with warm and bright colors, Um et al. (2012) demonstrated its beneficial impact on deep conceptual learning (as measured by transfer tests) and comprehension; whereas, Plass et al. (2014; Exp. 1) and Mayer & Estrella (2014) showed impact only on comprehension[2]. Plass et al. (2014; Exp. 2) demonstrated the beneficial impact of anthropomorphisms on comprehension, but only grey-scale anthropomorphisms had positive effect on transfer. Park et al. (2015) showed improvements in comprehension (but not in transfer) when a treatment with warm-bright-color anthropomorphisms was compared to a warm-bright-color, non-anthropomorphism baseline. At the same time, participants underwent a mood induction procedure prior to the learning experience. Both Um et al. (2012) and Plass et al. (2014) found that anthropomorphisms had positive impact on positive emotions/engagement. However, this was not so for Mayer and Estrella (2014) and Park and colleagues (2015). Generally, the impact of anthropomorphisms tends to be neutral to positive and somewhat unstable; effect sizes are in very small to large ranges (and also differing based on the control conditions).

This study's contribution is that it tests the effects of new "emotional" manipulations: additions of static black-and-white faces and two black-and-white funny elements (rather than manipulating color and/or adding expressive faces). We used three changes all at the same time, because the animation is dynamic (its visual appearance changes) and we wanted at least one enhanced element to be present on the screen at any moment.

Because sparse prior research demonstrated some benefits of anthropomorphic enhancements, we put forward the following hypotheses:

H1: Learners' state engagement induced by the learning, i.e., measured immediately after the learning, will be higher with the enhanced graphics.
H2: Learners' knowledge test scores will be higher with the enhanced graphics (for both retention and transfer).
H3: The effect of the graphics on learning results will be mediated by state engagement.

3. Method

3.1 Study Design

The study uses a between-subject quasi-experimental design with two conditions. Participants assigned into one of the conditions learnt about the process of how a biological treatment plant functions from about 6 min long animation. One condition used a neutral graphics (the N version, Figure 1) and the other emotional graphics (the E version; Figure 2). The animations did not differ in the expository instructions.

For the purpose of the present study, we recruited participants for the E condition only. In the present study, we compare learning outcomes and state engagement of these participants to learning outcomes and state engagement of participants recruited for the purpose of our *previous* study (Brom et al., submitted) into the N condition. (The previous study used more conditions but we use one of them for comparison in the present study as the N condition, as detailed below.) Therefore, participants were not randomly assigned into the E and N conditions. We instead recruited first the participants for the previous study (May 2014 - March 2015) and when that study ended, we started recruiting for the E condition (April 2015 - May 2015).

The primary dependent variables are generalized positive affect and flow levels (as proxies for state engagement), and retention and transfer test scores. The studies use also several auxiliary dependent variables, which we denote as "affective" variables: on generalized negative affect, on self-assessed learning, on the animation's usefulness, on the learner's level of interest and motivation, on the learner's perception of difficulty in learning from the materials, and on the animation's friendliness.

[2] Retention tests assess superficial learning: basically if the learners have memorized the content and are thus able to recall it. Transfer tests investigate deep understanding of the process/model learned by application of the knowledge in new situations (Mayer, 2009). Mayer and Estrella (2014) used classic retention and transfer tests. Um et al. (2012) and Plass et al. (2014) used transfer tests and comprehension tests. In our opinion, the latter were similar to retention tests.

3.2 Participants

Participants recruited for the E condition were 41 high school students (Mean age = 17.54; SD = 0.64) with intentionally diverse backgrounds. All students planned to pursue university studies (diverse study programs). There were 66% females. Students were recruited via an online server advertising short-term jobs for students. We invited them to participate in a 3 hour long experimental "workshop" for a compensation of 350 CZK (approximately 13 EUR). This workshop had two parts. The first part was the current experiment. The second part was unrelated to the present experiment (in the second part, we investigated how students learn from a certain educational game). The workshop was an extracurricular afternoon activity for the students.

Participants in the N version, who were originally recruited for the purpose of a different study (Brom et al., submitted), were recruited in the exactly same way as the E condition participants and participated under the same conditions. There were 37 participants in the N condition with the same background characteristics as the E condition participants (Mean age = 17.27; SD = 0.96; 54% females; diverse study background).

Eleven additional participants were excluded (N condition: 6; E condition: 5). The reasons for exclusion were: native language being neither Czech nor Slovak (Slovak is very close to the Czech language), being extremely tired at the beginning of the experiment, not answering questions on the knowledge tests.

3.3 Intervention

The intervention was a short, black-and-white animation explaining the process of biological wastewater treatment (see Brom et al., submitted for details). The animation consisted of 19 screens, each with a few lines of expository instructions placed at the bottom of the screen. The animation was controlled by learners using the "next" button. The average time to completion was around 6 minutes. The animation had 348 words.[3]

The visual appearance of the neutral version (Figure 1) was altered to produce the emotional version (Figure 2) in the following way:

a) two chemical elements were anthropomorphized by adding schematic human-like faces (the "anthropomorphism" design principle);
b) the fact that one of the chemicals is toxic for aquatic organisms and can cause mutations was highlighted by mutating the fish (the "fun" design principle);
c) a neutral appearance of garbage at the river bed was changed to a funny appearance (the "fun" design principle).

3.4 Paper materials

The paper materials consisted of a pretest with a background questionnaire, an initial motivation questionnaire, a prior and a post hoc inventory on generalized positive and negative affect, a post hoc flow questionnaire, a retention test, a transfer test, and a feedback questionnaire. All of these are further detailed below and in more detail in (Brom et al., submitted).

The background questionnaire yielded information about a participant's age, gender, study type, and native language. 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 scale), one question on the frequency of playing live action experiential/simulation games (a 5-point scale), and two questions on prior tiredness (a 6-point Likert scale). These questions were included primarily to verify that the groups were balanced with respect to these variables.

The pretest was based on Moreno and Mayer's self-assessing pretest (2000; Exp. 1, 2) and it was intended to measure participants' knowledge of chemistry and biological wastewater treatment. It contained ten questions (total scale 0-25). The question example is: "Please, make a check mark next to this sentence if it is TRUE in your case: 'I can thoroughly explain to a high school student how aerobic bacteria relate to wastewater treatment'".

[3] The original study (Brom et al., submitted, Exp. 4) compared two animations with the same graphics (i.e., the present neutral version) but with different texts; so-called "personalized" and "formal" versions. The "personalized" version of the texts is used in the present experiment (for both the E and N conditions).

The initial motivation questionnaire was a shortened version of the QCM inventory (Questionnaire on Current Motivation; Rheinberg, Vollmeyer, & Burns, 2001) and it was administered to assess participants' initial motivation to learn the given topic. Items with a 7-point Likert scale were organized into two factors: interest (5 items; Cronbach's α = .66) and anxiety (3 items; α = .73).

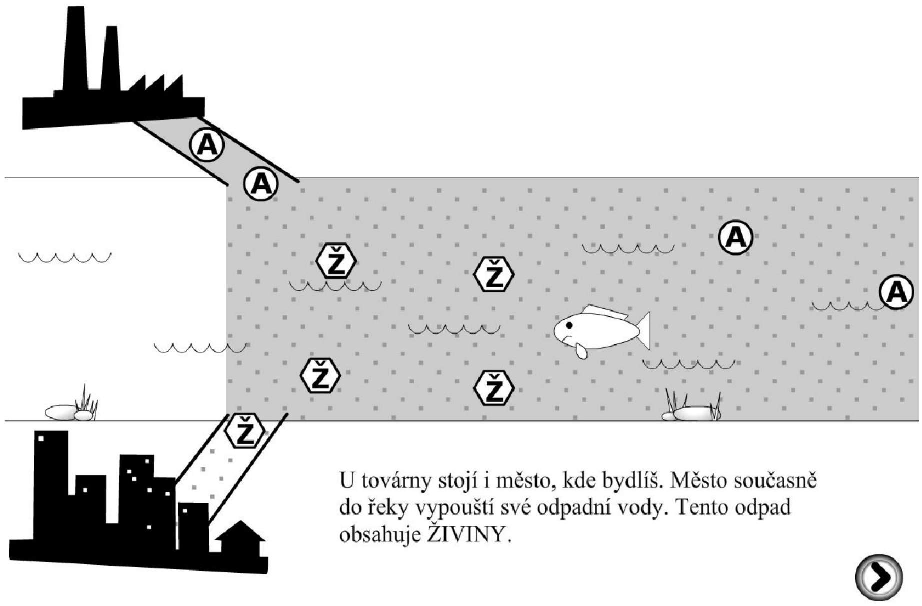

Figure 1: A screenshot of the neutral animation version.

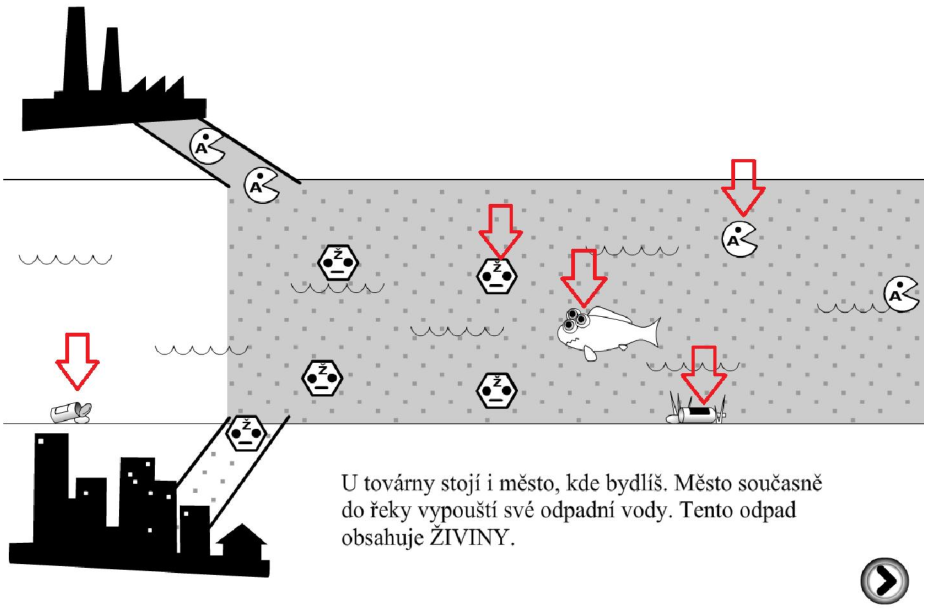

Figure 2: A screenshot of the emotional animation version. The arrows highlight the changes.

To assess flow levels, the Flow Short Scale inventory (Rheinberg, Vollmeyer, & Engeser, 2003) measuring induced level of flow with 10 items with a 7-point Likert scale was used (α = .80). As concerns the generalized positive and negative affect, we used the Positive and Negative Affect Schedule (PANAS) (Watson, Clark, & Tellegen, 1988). We measured affective state twice: immediately before and after the intervention. The schedule's positive dimension is denoted as PANAS+ and its negative dimension as PANAS-, the preintervention PANAS as PANAS1 and the post hoc PANAS as PANAS2. The tests' internal consistency was as follows: PANAS1+: α = .80; PANAS1-: α = .81; PANAS2+: α = .86; PANAS2-: α =.70.

Our retention test included the following open-ended question: "Based on the animation you just saw, describe in detail how biological wastewater treatment works". The transfer test included four open-ended questions. The question example is: "What would happen if a fungus first appeared in the treatment plant and then bacteria? Write down all consequences that come to mind based on the animation you saw today.". Participants could receive one point for each of 19 key idea-units in the retention test, or half a point for a partially correct idea-unit. These units reflected key steps in how a biological wastewater treatment plant functions as described in the animation. Likewise, participants received one point for every creative solution to a transfer test question, or half a point for a partially correct solution (open-ended scale). Solutions based only on prior knowledge were not rewarded.

The feedback questionnaire included the following questions with an 8-point Likert scale: two questions on self-assessed learning, one question on the animation's usefulness, two questions on the learner's level of interest, one question on the learner's level of motivation, one question on the learner's perception of difficulty in learning from the animation, and one question on the assessment of the animation's friendliness. The questionnaire also included eleven additional questions irrelevant for this study's purpose.

3.5 Interview

After interacting with the animation, participants in the E condition were briefly interviewed regarding the perception of general animation's usefulness, perception of graphics (Was or was it not funny?), and suitability of graphics for educational purposes (What do you think about the graphics? Did or did it not help in learning?). Participants were also asked what, if any, visuals they recalled during filling in of the tests and if that was helpful when answering the test questions.

3.6 Procedure

Participants were tested in groups of 1-9 per session, each seated at one computer with at least a 17" wide screen. After the introduction, participants filled in, at their own pace, pretests, background questionnaires, the questionnaire on initial motivation and the prior PANAS (note: participants had not yet seen animation graphics at that time). Afterwards, participants received the animation according to their condition; all participants started at once. Immediately after a participant finished, he/she received the Flow Short Scale followed by the post hoc PANAS and the feedback questionnaire. Afterwards, the retention test was administered, followed by the transfer test (the timing was strict for the two tests). Responses to each question were collected before the next question was distributed. Participants were then interviewed and thanked. Then, after a break, participants proceeded in the second part of the workshop (which is irrelevant for the present study).

4. Results and Discussion

Both groups were balanced with respect to the following variables: self-assessed knowledge of mathematics, frequency of videogame play, frequency of Larp participation, frequency of live action experiential/simulation game play, both factors of initial motivation, tiredness, both prior PANAS+ and PANAS-, and the average time during which participants completed the animation. We found significant differences as concerns pretest scores ($t(76) = 2.55$, $p = .013$, $d = 0.58$, 95% CI [0.12, 1.04]). Because this variable correlated with retention test scores ($r = .23$), though not with transfer test scores ($r = .07$), we took it as a covariate (and report the results both with and without taking the pretest as covariate).

Regarding qualitative data, during the interviews, participants commented on the graphics in the E version as being "appropriate for an educational animation", "well arranged" and/or "generally not distractive". The mutated fish was described as "funny" or "great", though sometimes as "distracting the learner from the learning"; the chemicals were described as "clearly understandable" and learners viewed them positively; and comments regarding the river bed were scarce. Some participants reported that the recall of particular visual elements helped them during filling in of the tests. Considering these points together, qualitative data indicate that the participants noticed the altered graphical elements, especially the fish and the chemicals' faces (which is a manipulation check); that our "anthropomorphic" manipulation might influence state engagement of learners; and the “funny” manipulation was noticeable regarding the fish, but less so regarding the river bed. Generally, the altered elements were not distractive, possibly with the exception of the fish for some participants. The case of fish probably exemplifies a trade-off between cognitive distraction and the ability to induce state engagement. The visual elements might serve as cues during filling in of the memory tests.

Quantitative data are presented in Tables 1 and 2. Regarding knowledge outcomes, there was no difference in transfer tests ($t(76) = 0.57$; $p = .572$; $d = 0.13$; 95% CI [-0.32, 0.58]; with correction for pretest: $F(1, 75) = 0.16$, $p = .687$, $\eta^2 = 0.002$, 95% CI [0.00, 0.03][4]), but there was a moderate difference regarding retention tests ($t(76) = 2.19$; $p = .031$; $d = 0.50$; 95% CI [0.04, 0.96) in favor of the E condition, which reduced to small, marginally significant difference after correcting for pretest scores ($F(1, 75) = 2.85$, $p = .095$, $\eta^2 = 0.038$, 95% CI [0.00, 0.18]).

Table 1: Means, SDs, and effect sizes for the retention and transfer tests.
[a]Uncorrected for pretest scores.

	E		N			
	M	*SD*	*M*	*SD*	*d*[a]	*95% CI*
Retention	11.71	3.07	10.07	3.50	0.50	[0.04, 0.96]
Transfer	6.29	2.23	5.98	2.49	0.13	[-0.32, 0.58]

Table 2: Means, SDs, and effect sizes for the affective variables.
[a]Scale 0 – 7. Higher values mean “more”.
[b]Scale 10 – 50. Higher values mean “more”.
[c]Scale 21 – 74 after the transformation through T-norms. Higher values mean “more”.

	E		N			
	M	*SD*	*M*	*SD*	*d*	*95% CI*
Learning[a]	5.17	0.91	5.09	1.12	0.08	[-0.38, 0.53]
Usefulness[a]	5.12	1.49	5.35	1.40	-0.16	[-0.61, 0.29]
Interest[a]	5.44	1.15	5.47	0.81	-0.03	[-0.49, 0.42]
Motivation[a]	6.27	1.00	6.24	0.95	0.03	[-0.43, 0.48]
Difficulty[a]	1.88	1.90	1.46	1.17	0.26	[-0.19, 0.72]
Friendliness[a]	6.27	0.87	6.27	0.65	0.00	[-0.45, 0.45]
PANAS1+[b]	27.74	5.50	28.28	6.21	-0.09	[-0.54, 0.36]
PANAS1-[b]	15.94	5.06	16.95	5.46	-0.19	[-0.64, 0.26]
PANAS2+[b]	30.23	6.54	30.91	7.15	-0.10	[-0.55, 0.35]
PANAS2-[b]	13.38	2.82	13.33	3.33	0.02	[-0.44, 0.47]
Flow[c]	54.10	6.36	55.26	8.17	-0.16	[-0.61, 0.29]
PANAS+ diff	2.71	5.00	2.59	4.28	0.03	[-0.43, 0.48]
PANAS- diff	-2.70	4.13	-3.81	3.90	0.27	[-0.19, 0.72]

Regarding self-reported variables, no difference was significant (all *p*s > .1; Table 2). Differences between the first and the second presentation of the PANAS were significant for both positive ($t(71) = 4.85$, $p < .001$, $d = 0.78$, 95% CI [0.45, 1.10]) and negative scale ($t(73) = -6.90$, $p < .001$, $d = -1.11$, 95% CI [-1.44, -0.77]). Flow levels were relatively high in absolute terms: they were in the same range as in the case of a college sample with the same nationality interacting with a 2-3 hour long brewery educational simulation and nearly a standard

[4] Confidence intervals in that case were computed using bootstrapping (N=1000).

deviation higher as in the case of high school students with the same nationality playing a complex, team-based educational game over 5 hours (Brom et al., submitted).

Considering all the data together, we conclude that both animation's versions induced state engagement, but to the same extent. Affectively, the "emotional" manipulations were too subtle to make noticeable changes (the possibility that effects of individual manipulations counterbalanced each other is improbable because there is no indication in the qualitative data that any manipulation influenced state engagement negatively). However, the funny elements may have a cognitive impact: the study's results are consistent with the idea that some of these elements might serve as cues for recalling the animation's content. Accordingly, the E version mildly enhanced superficial learning (as measured by retention tests), but the additional cues did not help in conceptual understanding of the whole process (which is not surprising). It is thus possible that certain supposedly emotional manipulations have actually direct cognitive impact, unmediated via increased state engagement. The study of Mayer and Estrella (2014) and to a lesser extent also the studies of Plass et al. (2014; Exp. 1) and Park et al. (2015) showed a larger impact of anthropomorphisms on retention than on transfer. Only the original study of Um et al. (2012) clearly demonstrated the impact on both types of knowledge outcomes. Our study also highlighted the possibility of distraction by overly funny elements (the fish), but in our particular case, the distraction was probably small or outweighed by positives of the element's abilities to serve as a cue – because no decrease in knowledge outcomes was detected.

We can conclude that Hypothesis 1 has not been supported by the data. Hypothesis 2 was partly supported: regarding retention tests (with small effect size). Because there was no effect of the enhanced graphical appearance on state engagement, we cannot investigate its role in mediating the influence of the enhanced graphics on learning outcomes (i.e., Hypothesis 3).

5. Conclusion and Limitations

This study investigated the effects of funny graphics and static anthropomorphisms on state engagement and learning outcomes in the context of a 6-minute-long, black-and-white, instructional animation. In general, the graphics in the "emotionally" enhanced version were viewed as positive, funny, accurate, and not distracting (perhaps with the exception of one element – the funny, mutated fish – for some participants). Yet the manipulation of the graphics was too subtle to yield measurable between-group differences regarding state engagement (indexed by generalized positive affect and flow level) and regarding deep conceptual learning (assessed by transfer tests). Only a small effect was found as concerns superficial learning (assessed by retention tests). At the same time, the learning experience induced state engagement in both groups.

Our finding is similar to what has been reported by Mayer and Estrella (2014) (anthropomorphisms in combination with warm, bright colors vs. schematic, black-and-white graphics), Plass et al., (2014; Exp. 1) (warm-bright-color anthropomorphisms vs. grey-scale, non-anthropomorphic graphics), Park et al. (2015) (warm-bright-color anthropomorphisms vs. warm, bright colors with round shapes, but no anthropomorphisms), and Snetsinger and Grabowski (1993) (funny appearance vs. neutral appearance). In fact, to our knowledge, only Um et al. (2012) demonstrated clear effects of emotional enhancements, warm-bright-color anthropomorphisms in their case, both in affective an cognitive terms, and on both comprehension and deep conceptual learning. In addition, Plass and colleagues (2014; Exp. 2) showed a positive effect of anthropomorphisms on comprehension, but only grey-scale (and not warm-bright-color) anthropomorphisms had positive effect on transfer.

In terms of our explanatory framework, CATLM, state engagement of the participants in the "enhanced" graphics group probably did not increase notably. Therefore, allocated cognitive resources did not increase notably, and, as a consequence, learning was not much enhanced: especially not deep, conceptual learning. Because we made three graphical alterations (rather than just one), it is, in theory, possible that a positive effect of a particular element was counterbalanced by a negative effect of a different element. However, there is no support for this explanation in the qualitative data. Therefore, we opt for the former, more parsimonious, explanation.

A primary practical implication of this study is that, as concerns instructional animations, adding black-and-white, static faces to otherwise non-human graphical elements and changing a few black-and-white elements to appear slightly more funny is probably not a big deal for high school (~ 17 years old) learners. The open question is the following: what would be a big deal? Given the importance of affective states in academic and achievement contexts, research to answer this question is much needed. One option is addition of color. However, as the results of Heidig and colleagues (2015) showed, color alone might also make no difference (i.e., if color is used to increase appeal, not for signaling purposes) (see also Plass et al., 2014, Exp. 2, for somewhat mixed findings). A combination of color and other features may be an option, as suggested by the work of Um and colleagues (2012), but this finding was only partially replicated (Plass et al., 2014, Exp. 1; i.e., for positive affect and comprehension but not for transfer; Mayer & Estrella, 2014; i.e., for comprehension only). One study demonstrated the highest effects on transfer for gray-scale rather than warm-bright-color anthropomorphisms (Plass et al., 2014, Exp. 2). More expressive anthropomorphisms may also be an option, but researchers will probably have to start looking beyond anthropomorphisms and funny appearances when searching for robust emotional design principles, for instance, on adding sophisticated game-based forms of interactivity (cf. Plass & Kaplan, 2015, pp. 139-142).

This also has a methodological implication. Studies investigating what role state engagement plays in enhancing learning would need to find more robust manipulations than those discussed here.

A theoretical implication of this study is that the impact of a graphical manipulation on learning outcomes, even if it is supposed to be mediated via state engagement, may be of cognitive nature. The only notable effect of our "enhanced" graphics was on retention tests scores, i.e., on rote memorizing what the materials were about. Better cues may help in recalling this content, and our qualitative data suggested that some of the "enhanced" elements actually might serve as such cues (whereas, these cues might not aid in deep, conceptual learning). In the future, it would also be useful to consider, in emotional design studies, the possibility that the manipulation in question may have a direct (i.e., unmediated via increased state engagement) effect on learning. Park and colleagues (2015) showed that expressive anthropomorphisms can grab learners' attention (as demonstrated by eye tracking data) without changing state engagement; but with a facilitative effect on comprehension, when the learning experience is preceded by a mood induction procedure.

Finally, this study is not without limitations, and so the above interpretations should be treated cautiously. First, this was a quasi-experimental study only. Participants were not randomly assigned to conditions and we cannot exclude the possibility that slightly different participant groups arrived to our conditions (the recruitment process was the same, but the two groups were investigated at different times of the year). Second, the neutral graphics group came from a different experiment. Thus, we did not have the opportunity to ask the N group participants in the interview the same questions that we asked the E group learners. For instance, the E group participants considered the "enhanced" graphics funny. Is it possible that the N group participants would also have considered the neutral graphical images funny, even if these images were not supposed to be? Third, even if a sample size of around 30 participants per cell is not atypical in the context of educational intervention research, the present sample size, i.e., 41 + 37, is actually relatively small (as also apparent from large confidence intervals for effect sizes). The study could thus miss a true difference with a small effect size.

Acknowledgements

This research was partially funded by the research project nr. 15-14715S supported by the Czech Grant Science Foundation (GA ČR). Tereza Hannemann was partially supported by Charles University Grant Agency (GA UK), research project nr. 1315. Filip Děchtěrenko was partially supported by SVV project number 260 333.

References

Brom, C., Hannemann, T., Stárková, T., Bromová, E. and Děchtěrenko, F. (submitted as of 13 June 2016) "The Role of Cultural Background in the Personalization Principle: Five Experiments with Czech Learners"

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

D'Mello, S. (2013) "A selective meta-analysis on the relative incidence of discrete affective states during learning with technology", *Journal of Educational Psychology*, Vol. 105, No. 4, pp 1082-1099.

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*, Vol. 44, pp 81-95.

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

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

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

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

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

Moreno, R., & Mayer, R. E. (2000) "Engaging students in active learning: The case for personalized multimedia messages", *Journal of Educational Psychology*, Vol. 92, No. 4, pp 724-733.

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

Pekrun, R. (2006) "The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice", *Educational Psychology Review*, Vol. 18, pp 315–341.

Plass, J. L., Heidig, S., Hayward, E. O., Homer, B. D. and Um, E. (2014) "Emotional design in multimedia learning: Effects of shape and color on affect and learning", *Learning and Instruction*, Vol. 29, pp 128-140.

Plass, J. L. and Kaplan, U. (2015) "Emotional Design in Digital Media for Learning" in *Emotions, Technology, Design, and Learning,* pp 131–161.

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,* Vol. 47, pp 57-66.

Rheinberg, F., Vollmeyer, R., & Engeser, S. (2003) "Die Erfassung des Flow-Erlebens" [in German] in *Diagnostik von Motivation und Selbstkonzept*, Hogrefe, pp 261-279.

Snetsinger, W. and Grabowski, B. (1993) "Use of Humorous Visuals To Enhance Computer-Based-Instruction" in *Visual Literacy in the Digital Age: Selected Readings from the Annual Conference of the International Visual Literacy Association* (25th, Rochester, New York, October 13-17, 1993).

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

Um, E. R., Plass, J. L., Hayward, E. O. and Homer, B. D. (2012) "Emotional Design in Multimedia Learning", *Journal of Educational Psychology*, Vol. 104, No. 2, pp 485-498.

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 Psycholog*y, Vol. 54, No. 6, pp 1063-1070.

Wouters, P., van Nimwegen, C., van Oostendorp, H. and van der Spek, E. D. (2013) "A Meta-Analysis of the Cognitive and Motivational Effects of Serious Games", *Journal of Educational Psychology*, Vol. 105, No. 2, pp 249-265.