



The role of process information in narrations while learning with animations and static pictures



Ferdinand Stebner ^{a,*}, Tim Kühl ^b, Tim N. Höffler ^c, Joachim Wirth ^a, Paul Ayres ^d

^a Ruhr-University Bochum, Department of Research on Learning and Instruction, Postfach 102148, 44721, Bochum, Germany

^b University of Mannheim, Psychology of Education, Postfach 103462, 68131, Mannheim, Germany

^c IPN – Leibniz-Institute for Science and Mathematics Education at the University of Kiel, Olshausenstr. 62, 24098, Kiel, Germany

^d School of Education, University of New South Wales, Sydney, NSW, 2052, Australia

ARTICLE INFO

Article history:

Received 23 May 2016

Received in revised form 23 September 2016

Accepted 1 November 2016

Available online 2 November 2016

Keywords:

Multimedia

Animation

Static pictures

Narration

Cognitive load

ABSTRACT

The role of process information in annotating narrations used for learning with animations compared to static pictures is examined. In two experiments, seventh and eighth graders from German high schools were randomly assigned to learning environments which differed in the combination of visualization (no visualization vs. static pictures vs. animation) and type of narration (no narration vs. non-process narration vs. process narration). Results revealed that visualizations were necessary for this kind of instructional material to gain a deeper understanding. Moreover, the results consistently show a significant superiority of animations over static pictures. Concerning narrations, results display a significant superiority of process descriptions only in Experiment 1. Contrary to prior assumptions, the interaction of specific information in narrations with the type of visualizations was not significant.

© 2016 Elsevier Ltd. All rights reserved.

1. Theoretical background

Technological development enables education systems to create and combine enormous amounts of different learning materials. However, there is still a large gap between advanced technology and our understanding of how humans can best learn with this technology (Chandler, 2004). The demands of multimedia learning settings are described in Mayer's Cognitive Theory of Multimedia Learning (CTML; Mayer, 2001, 2006, 2009). Based on the assumptions that the learners' working memory is limited in capacity (Baddeley, 1986; Sweller, 1999) and that it is structured in an auditory and a visual channel (Paivio, 1986), the learner has to select and organize new verbal and visual information to build single models, and integrate existing knowledge to build the final mental model.

A further theory describing cognitive processes during learning is the Cognitive Load Theory (CLT; Sweller, 1988, 2010). According to CLT, cognitive load is caused by the complexity of the learning content (intrinsic load) as well as by the design of the learning material (extrinsic load). To prevent a cognitive overload (and inefficient learning), learning environments should be designed using considerations of cognitive load and human cognitive architecture in general (Sweller, Ayres, & Kalyuga, 2011).

* Corresponding author.

E-mail address: ferdinand.stebner@rub.de (F. Stebner).

In multimedia contexts, learning with dynamic visualizations like animations has become a topic of major interest in the last decade (cf. Höffler, Schmeck, & Opfermann, 2013; Lowe & Schnotz, 2014). In comparison to static pictures, the term ‘animation’ can refer to any display element that *changes* its attributes over time (Schnotz & Lowe, 2008). It can be defined as a series of rapidly changing computer screen displays suggesting movement to the viewer (Rieber & Kini, 1991). When comparing the effectiveness of animations and static pictures, the question whether one or the other is better suited for learning in general may not be productive and often leads to inconclusive results (cf. Berney & Bétrancourt, 2016; Höffler & Leutner, 2007; Tversky, Morrison-Bauer, & Bétrancourt, 2002). A more helpful focus is to investigate the conditions for which animations might be more appropriate than static pictures – and vice versa (Bétrancourt, 2005; Rieber, 1990).

Thus, further theories or models might be needed to describe how (dynamic) mental models are built up when learning with static or dynamic media; in this case, Narayanan and Hegarty (1998), for instance, developed a model of how dynamic systems are learned. The “runnable” mental model is a result from the process of decomposing a dynamic system into simpler components, retrieving relevant background knowledge about these components and mentally encoding the relations between components to construct a static mental model. This static model then needs to be mentally animated (Narayanan & Hegarty, 2002). It becomes obvious that the process of building a dynamic mental model might be significantly influenced by the characteristics of each medium and especially by the combination of different media. Imagine a process is to be learned: Whereas in case of animations, the process is already dynamically presented and “only” needs to be selected, the learner has to animate static information autonomously when using static pictures. Adding narrations which explain the process might then support the learning process differently due to the fact that the visualizations show the process differently.

When animations and static pictures are accompanied by textual information, which is usually the case in multimedia learning (cf. Mayer, 2009), the role of the textual information in supporting the different types of visualizations is still unclear, despite considerable research. This issue will be addressed in the following sections.

1.1. Learning with animations and static pictures

Empirical studies on learning with animations and static pictures have yielded somewhat inconclusive results: On the one hand, meta-analyses revealed a medium advantage of animations over static pictures (Berney & Bétrancourt, 2016; Höffler & Leutner, 2007). On the other hand, there are several studies showing no advantage or even a disadvantage for learning with animations (e.g., Castro-Alonso, Ayres, & Paas, 2014; Hegarty, Kriz, & Cate, 2003; Mayer, Hegarty, Mayer, & Campbell, 2005; Tversky et al., 2002). Such mixed results indicate that it is necessary to take a deeper look and determine exactly when learning with animations is most promising.

Since spatial changes of elements are depicted in animations – in comparison to static pictures – directly, they are supposed to be particularly well suited for situations in which processes are to be learned (e.g., chemical processes). In such instances, animations might be expected to not unduly increase cognitive load (see CLT; Sweller et al., 2011) because of their relevance, and therefore support learners in constructing a dynamic mental model of the content (Hegarty et al., 2003; Schnotz & Lowe, 2008). However, a noted drawback of animations is their transient nature (see Ayres & Paas, 2007; Wong, Leahy, Marcus, & Sweller, 2012). Because dynamic representations flow forward frame-by-frame, important information can be lost from view before the learner has time to adequately select and process this information. This can raise cognitive load as the learner must temporarily store previously viewed information, while processing and linking it with new information. If cognitive load exceeds the limit of working memory, a cognitive overload might diminish learning success. One way of counteracting the demands imposed by the transient nature of animations is to implement interactive elements, so that learners have the chance to stop or replay the animation in instances of high cognitive load (e.g., Boucheix, 2008; Höffler & Schwartz, 2011; Lowe, 2008; Schwan & Riempp, 2004). Another way of counteracting is to use annotating narrations as a second source of information which might highlight relevant visual aspects (Roscoe, Jacovina, Harry, Russell, & McNamara, 2015).

1.2. The role of annotating narrations

Although the understanding of how humans process only a single medium like animations is still in its infancy, there is a big necessity to explore the relation between both visualization and narration (Plötzner & Lowe, 2004; Schmidt-Weigand & Scheiter, 2011). A narration is a spoken version of text, which can describe, supplement, and highlight what can be seen in the visualization (Tversky et al., 2002). In comparison to written text, the learner processes this information in the auditory channel, which might be crucial in terms of learning success when the learner is simultaneously watching visualizations (cf. modality effect; Low & Sweller, 2014). In multimedia learning, however, it is not sufficient to consider each type of medium in isolation because media interact with one another in a form of ‘representational chemistry’ (Ainsworth, 2006).

In the context of multimedia learning environments, which are by definition comprised of text and visualizations (Mayer, 2009), the so-called multimedia principle (Mayer, 2009) states that visualizations added to text should *enhance* learning; this has been investigated extensively (e.g., Mayer & Anderson, 1991, 1992; Mayer, 1989; Moreno & Mayer, 2002) and can thus be considered as well-established. If text by itself was sufficient for understanding and led to the same learning outcome, additional visualizations would not be necessary at all. This is the case when visualizations serve merely for decorative purposes (see Höffler & Leutner, 2007; Mayer, 2009). Along the same lines, it has been shown that visualizations without text can be sufficient for learners if the text contains redundant material that is also represented in the visualizations (see

Chandler & Sweller, 1991). What both of these findings suggest is that it is important to eliminate redundant materials, but also that single modality materials serve as important control conditions when investigating sophisticated multimedia effects. The redundancy effect, however, is interpreted differently: On the one hand, Kalyuga, Chandler, and Sweller (2000) state that “the redundancy effect generally occurs under conditions in which different sources of information are intelligible in isolation and in which each source provides similar information but in a different form” (p. 127). On the other hand, Mayer uses the term redundancy in a more restricted sense by saying that “people learn better from animation and narration than from animation, narration, and on-screen text” (Mayer, 2006, p. 376), where on-screen text is the redundant information. In a meta-analysis, Adesope and Nesbit (2012) examined verbal redundancy in multimedia learning environments. They could show that outcomes comparing spoken-plus-written and written-only presentations did not differ, but spoken-plus-written was superior to spoken-only conditions. This effect was moderated by learners’ prior knowledge, pacing of presentation and inclusion of animations or diagrams. To sum up those results: Redundancy was only helpful when learners had low prior knowledge, when the learning material was system-paced and when no pictures were added (Adesope & Nesbit, 2012).

Even though texts are an essential part of learning with animations and static pictures, there is surprisingly little research considering its potential moderating impact. This type of research is difficult to design because of the need to provide equivalent static and animated conditions (see Tversky et al., 2002), and associated difficulties synchronizing spoken and written text with static and moving pictures. Amongst the few, Köhl, Scheiter, Gerjets, and Edelmann (2011) investigated if the modality of the text (i.e. spoken vs. written text) moderates learning outcomes depending on whether students learned with animations or static pictures. Köhl et al. (2011) found an instructional advantage of animations over static pictures as well as an instructional advantage of spoken over written text (cf. modality effect; Low & Sweller, 2014), but text modality did not act as a moderator.

Catrambone and Seay (2002) examined the impact of text coherence (poor vs. good) on learning with animations compared to static pictures in the domain of algorithms. Animations turned out to be superior to static pictures for poorer texts exhibiting a compensating effect, but there were no differences between the types of visualizations with respect to good texts.

Similar results were found by Schmidt-Weigand and Scheiter (2011) when investigating the role of an animation compared to no visualization while learning with different types of texts. Concerning the amount of spatial information presented by the text, animations were especially helpful in the text condition with low spatial information.

Generally, most studies which have compared the effectiveness of animations and static pictures have also used annotating texts (cf. Höffler & Leutner, 2007). Furthermore, if a process is to be learned, most of these texts used in empirical studies also contain process information (e.g., Hegarty, 1992; Höffler & Leutner, 2011; Large, Beheshti, Breuleux, & Renaud, 1996; Lee & Shin, 2011; Plötzner, Bodemer, & Neudert, 2008). Whereas process information provided by annotating narrations could be considered complementary to the information provided by static pictures (“you hear what you do not see”), it might be redundant when provided alongside animations (“you hear what you see”). Such redundancy might increase a learner’s cognitive load because of the additional processing of the redundant materials, which are unnecessary for learning (cf. redundancy effect; Kalyuga et al., 2000; Kalyuga & Sweller, 2014; Sweller et al., 2011). In this case, it would be more beneficial to exclude process information from accompanying narrations for animations, but include it for static pictures to facilitate a better understanding of the process.

In a nutshell, while for learning with animations a text with no process information might be more advantageous than one with process information, for learning with static pictures a text with no process information might be more disadvantageous. Hence, the process information in texts might function as a moderator when learning with animations and static pictures.

1.3. Research question and hypotheses

As argued above, previous research investigating main effects of animations and static pictures mostly neglected the impact of the kind of information given by annotating narrations. However, because of redundancy effects, it could be expected that process information provided through an extensive verbal description of a process affects learning with animations and static pictures in a different way.

To examine this assumption, nine different learning conditions consistent with a 3x3-design with visualization (no visualization vs. static pictures vs. animation) and narration (no narration vs. non-process narration vs. process narration) were constructed (for further more detailed explanations see section 2.1 *Instructional Materials*). Visualizations and narrations described the chemical processes during washing laundry. Previous studies using a similar version of our learning material did confirm an advantage of animations over static pictures whereby both visualizations were accompanied with the same narration which included process descriptions (cf. Höffler & Leutner, 2011).

The design described above made it possible to test a number of predictions. Firstly, it was hypothesized that visualizations and narrations together were necessary to understand the instructional material. According to the multimedia effect (e.g., Mayer, 2009), the participants in the experimental conditions with visualizations and narration should outperform those in the narration-only conditions (Hypothesis 1a) as well as the visualization-only conditions (Hypothesis 1b). Furthermore, the multimedia effect should also be mirrored by the participants’ cognitive load. Therefore, cognitive load should be higher for those in the narration-only conditions (Hypothesis 1c) as well as in the visualization-only conditions (Hypothesis 1d) as these conditions do not have the benefit of dual coding advantages in forming mental representations of the concepts.

For the content used in the current study, it has already been shown that animations were instructionally superior to static pictures (Höffler & Leutner, 2011), and therefore we expected to replicate these findings by finding higher learning outcomes (Hypothesis 2a) as well as lower cognitive load (Hypothesis 2b) for participants in the animation conditions.

Thirdly, based on the theoretical derivation (see section 1.2), it was hypothesized that the kind of textual information (process/non process) of the accompanying narration would moderate learning with animations and static pictures (Hypothesis 3). More precisely, learning with animations should lead to higher learning outcomes when the accompanying text contains less instead of more process information, whereas learning with static pictures would be more successful when the accompanying text contains more instead of less process information.

2. Experiment 1

2.1. Method

2.1.1. Participants and design

We calculated the required sample size for the 3x3-Between Subject Design with the software tool G*Power 3.1 (Faul, Erdfelder, Buchner, & Lang, 2009). Assuming a medium effect size $f = 0.25$, a Type I error of 0.05 and a Type II error of 0.20, the minimum required total sample size was 196 participants. Using even more participants is not harmful, but reasonable as long as more participants than required can be easily acquired – as was the case for the current experiment.

Finally, 283 seventh and eighth graders from four German secondary schools (46.6% female; $M = 13.48$ years, $SD = 0.67$, eleven different school classes) participated in the study and were randomly assigned to one of nine conditions consistent with the 3x3-design, with visualization (no visualization vs. static pictures vs. animation) and narration (no narration vs. non-process narration vs. process narration) as independent variables. Due to the German school system, learning about surfactants is content of ninth grader's curriculum. Therefore, seventh and eighth graders were asked to take part in our study to ensure participants' low prior knowledge. The students indeed only achieved on average $M = 1.46$ ($SD = 0.94$) points in the prior knowledge test, whereas 18 points were attainable.

The numbers of participants for each condition are presented in Table 1. The study was conducted in the students' classrooms. The size of classes differed between 20 and 30 students.

2.1.2. Instructional materials

The computerized instructional material was adapted from Höffler and Schwartz (2011) and consisted of six html-pages. The topic was the role of surfactants for the washing process describing how they are able to chemically remove dirt from clothes. All of the nine conditions consisted of the same introductory “welcome site”, where the use of the learning environment was explained. Also, the introductory material on general subject-matter aspects was identical and distributed on four pages: 1) Definition of surfactants, 2) Characteristics of surfactants in the water, 3) Surface tension, and 4) Decrease of surface tension. This introductory material was presented as written text with static pictures. Learners could navigate through the introductory material in a self-paced manner (i.e., back and forth buttons).

Table 1
Means (and SD) as a function of text version and visualization format (Experiment 1).

Text version	No narration			Non-process description			Process description		
	No	Static pictures	Anima-tion	No	Static pictures	Anima-tion	No	Static pictures	Anima-tion
<i>n</i> after exclusion	27	37	26	23	26	26	26	30	31
<i>n</i> before exclusion	27	40	29	28	30	28	30	31	40
Control variables									
Prior knowledge	1.11 (0.64)	1.24 (0.64)	1.38 (0.50)	1.26 (0.62)	1.04 (0.60)	1.08 (0.48)	1.31 (0.62)	1.20 (0.62)	1.52 (0.57)
Spatial abilities	5.07 (2.22)	4.95 (1.84)	5.89 (2.12)	5.83 (1.85)	5.54 (1.58)	5.54 (1.84)	5.15 (2.62)	5.67 (2.06)	5.65 (2.12)
Cognitive load									
Difficulty learning	5.73 (1.22)	5.53 (1.72)	5.15 (1.64)	5.65 (1.97)	5.85 (1.67)	5.31 (1.57)	5.54 (1.24)	5.70 (1.49)	4.94 (1.73)
Difficulty testing	7.00 (1.35)	6.58 (1.24)	5.71 (1.29)	6.59 (1.42)	6.31 (1.41)	6.12 (1.39)	6.44 (1.43)	6.20 (1.19)	5.44 (1.49)
Effort learning	5.41 (1.93)	5.58 (1.56)	5.60 (1.73)	6.26 (1.60)	5.58 (1.81)	5.42 (1.68)	5.36 (1.63)	5.80 (1.08)	5.23 (1.52)
Effort	5.52 (1.95)	5.58 (1.57)	5.29 (1.66)	5.59 (1.59)	5.54 (1.59)	5.96 (1.62)	5.69 (1.53)	5.64 (0.81)	5.42 (1.13)
Learning outcomes									
Introduction	2.92 (2.59)	3.60 (2.53)	5.27 (2.96)	3.44 (2.09)	4.04 (2.84)	3.35 (3.02)	2.88 (2.51)	3.76 (2.72)	4.13 (2.80)
Compre-hension	5.56 (4.00)	11.19 (5.36)	15.27 (5.49)	7.00 (4.31)	14.69 (6.52)	14.85 (6.06)	8.46 (5.76)	15.80 (6.37)	17.97 (6.40)

The subsequent html-page addressed the crucial chemical process under the headline “Removing dirt from the fibre” and was subject to experimental manipulations according to the nine conditions. With respect to the three different versions of *visualizations*, there was one version that had *no visualization* at all, but simply displayed a green background. The condition of *static pictures* consisted of four static key frames of the washing process. These simultaneously presented frames were taken from the animation and arranged in a 2x2-matrix (see Fig. 1). Each frame was depicted in the maximum possible size that still fitted on the screen. The *animation* depicted the washing process, explicitly showing spatial and temporal changes. It lasted for 73 s, which was also the duration of the narration. The size of the animation was the same as the 2x2-matrix, meaning that the animation was larger than each of the single key frames. We refrained from artificially decreasing the size of the animation.

With regard to the three different versions of the *narration*, there was one zero version that had *no narration* (or text) at all. The *non-process narration* (see appendix) explained how surfactants remove dirt from clothes and consisted of 142 words. However, it contained no explicit descriptions of movement (spatial changes) or chronological order (temporal changes). The *process narration* (see appendix) also explained how surfactants remove dirt from clothes and consisted of 144 words. In contrast to the other description, the process version contained explicit descriptions of movement (e.g., the hydrophobic part of the surfactants turns/is oriented away from the water [towards the fibre]; the negative particles repel each other) or descriptions of chronological order (e.g., at the beginning; thereafter; at the end). Both narrations were spoken by the same female speaker and lasted for 73 seconds.

For all of the nine conditions, there was a back-as well as an end-button for this last site. Also, there was a play-button, which activated the respective narration of the accompanying animation/static pictures in case there was one. No further options for interactivity were implemented (i.e., the narration and or animation could not be paused, rewound, forwarded, etc.), but after the narration/animation was finished, learners could press the play-button again and do so as often as they liked within ten minutes learning time. Since for the condition “no narration and no visualization” nothing would have been shown or played, it was decided to show a narrated animation of the same duration about another topic (usage of hydro-thermal energy). For the condition “no narration and static pictures”, there was no play-button, since there was no feature that could have been played (learners were solely able to watch the four static pictures). When pressing the end-button on

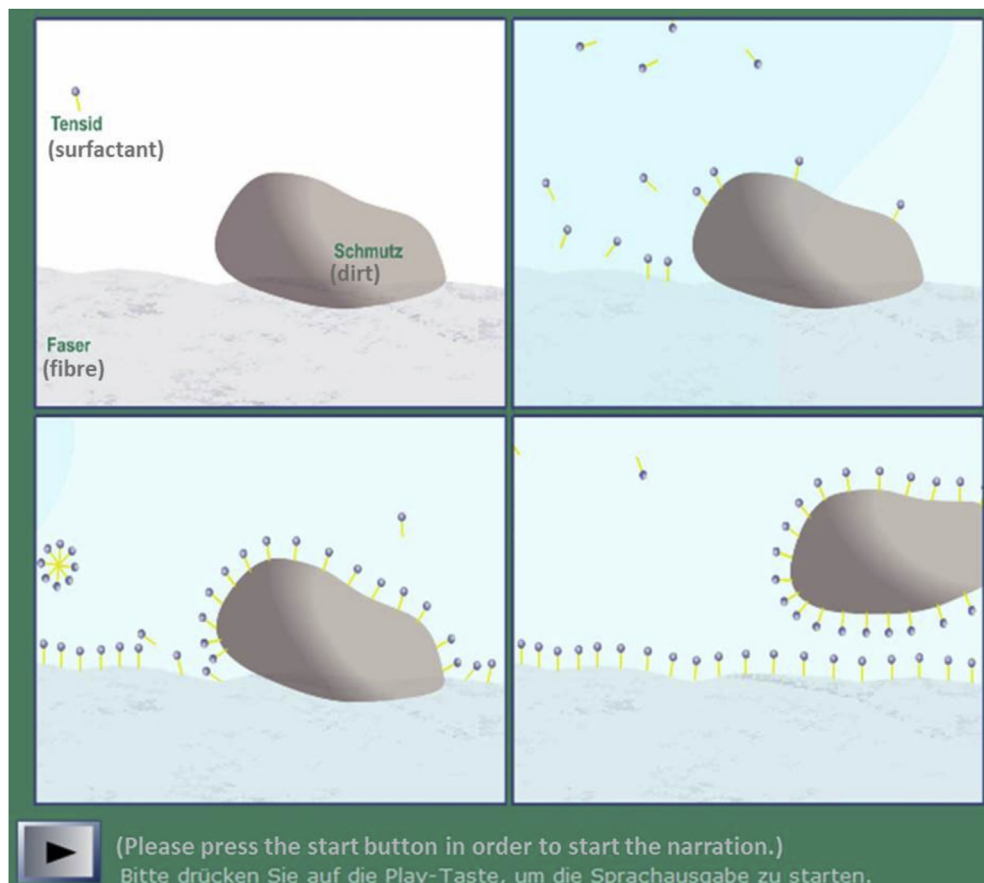


Fig. 1. Four key frames (presented simultaneously) from the static-picture condition.

this page, a final page showed up, instructing learners to signal to the assessor that they were finished with the learning phase. There was a maximum time limit of ten minutes for the learning phase. The students learned in a self-regulated manner, therefore, the exact amount of learning time could have varied due to individual circumstances. The exact amount of learning time was, however, not measured.

2.1.3. Measures

The measures consisted of a prior knowledge test, a spatial ability test, and subjective ratings of cognitive load, as well as a knowledge test to measure learning outcomes. All measures were paper-based and delivered by the assessor.

Prior knowledge test. The prior knowledge test contained three questions about 1) the process of how a detergent cleans laundry, 2) what the term *surface tension* means, and 3) what the term *surfactant* means. Each of the three questions was posed in an open format and participants were given eight minutes to write down the answers. For the scoring, a predefined list of correct answers was used and for each correct aspect, one point was given. Overall, learners could theoretically have achieved a maximum of 18 points.

Spatial ability test. Spatial abilities were assessed by means of a shortened version of the Paper Folding Test (PFT; Ekstrom, French, Harman, & Dermen, 1976). The shortened version of the PFT consisted of ten items which had to be solved within three minutes. For each correct answer one point was given, resulting in a minimum of 0 and a maximum of 10 points. This test was included because some studies about animations have shown that spatial ability can moderate their effectiveness (cf. Höffler, 2010).

Subjective ratings of cognitive load. To assess cognitive load subjective ratings of two items were used: The first item asked for perceived difficulty (“How easy or difficult was it for you to work on this task?”; Kalyuga, Chandler, & Sweller, 1999) and had to be rated on a 9-point Likert scale ranging from 1 (very, very easy) to 9 (very, very difficult). The second item asked for the invested mental effort (“How much mental effort did you invest to work on this task?”; Paas, 1992) and also had to be rated on a 9-point Likert scale ranging from 1 (very, very low) to 9 (very, very high). These items were used three times (after learning, after the first half as well as after the second half of the post knowledge test). Therefore, four different constructs are reported: the load (perceived difficulty and mental effort, respectively) experienced during learning and the average load (again, perceived difficulty and mental effort, respectively) experienced during testing.

Post knowledge test. Learning outcomes were measured in the post knowledge test which consisted of 18 questions in total, containing multiple-choice questions, questions in an open format, as well as pictorial tasks. For the scoring of the open questions, a predefined list of correct answers was used and for each correct aspect, one point was given. The open questions were scored by two independent raters (Cohen’s kappa = 0.85). Four of these 18 questions addressed contents of the introductory material (which was identical for all conditions) and the number of points was summed up (Introduction-Score). The remaining 14 questions addressed content that was presented in the experimental phase and was thus subject to experimental variations. Learners’ scores on these questions were totaled to give a Comprehension-Score. Learners were given 25 min in total to solve the tasks.

2.1.4. Procedure

When the participants entered the classroom, they had to choose their own working space endowed with a table, a chair, a laptop and a computer mouse to control the laptop. Each laptop was equipped with one of the nine different learning environments. However, from the outside they all looked exactly the same. Beforehand, the laptops were randomly stationed on one of the tables. Therefore, neither the assessor nor the participants knew with what kind of learning environment the participants would have to learn. Furthermore, the participants did not know that the other participants might have learned with different material.

After the welcoming and the instruction, the prior knowledge test was administered. After eight minutes, participants started working simultaneously on the spatial ability test. Then, participants started the learning environment. After the time limit of ten minutes, participants rated the two cognitive load statements, before they went on to work on the knowledge test for 25 min. At half way and after the knowledge test, participants again rated their experienced cognitive load. The whole procedure took about 90 min.

2.1.5. Statistical methods

Data was analysed within the framework of the General Linear Model (Horton, 1978) using analyses of variance (ANOVA). A calculation of intraclass correlation coefficients (ICC) revealed that the data was not hierarchically clustered regarding the different school classes participants were from (ICCs < 0.08). Thus, multi-level analyses were not necessary. Regarding the main outcome variable, we also checked for assumptions of ANOVA – homogeneity of variance, normality, and independence of observations. In one subgroup, the assumption of normality was violated (measured with a Kolmogorov-Smirnov test). However, ANOVAs are considered as fairly robust against violations of normality (Glass, Peckham, & Sanders, 1972). Therefore, we refrained from using non-parametrical methods due to the fact that we needed to test two-factorial models.

2.2. Results of experiment 1

Means and standard deviations are reported in Table 1. For all statistical tests a significance level of 0.05 was applied. For ANOVAs and *t*-tests, partial eta-squared (η^2_p) is reported as a measure of effect size.

2.2.1. Control variables

To examine whether participants had similar cognitive prerequisites across all experimental conditions, one-factorial ANOVAs were conducted to test for levels of prior knowledge and spatial abilities.

With regard to spatial ability, a one-factorial ANOVA showed no differences with respect to participants' conditions, $F(8, 274) = 1.29, p = 0.25$. In contrast, a one-factorial ANOVA revealed that participants' prior knowledge could not be considered equal across conditions, $F(8, 274) = 3.11, p = 0.002, \eta^2_p = 0.083$. These differences were too large to be statistically controlled for ($\eta^2_p > 0.06$; cf. Slavin, 1986). Thus, all those participants who had too much prior knowledge, using a cut-off value of 3 points or more, were excluded (31 participants; 10.9% of all participants). The numbers of participants for each condition before and after the exclusion are presented in Table 1. A further ANOVA without the excluded participants showed that there were no longer significant differences between conditions regarding both learners' prior knowledge ($F(8, 243) = 1.89, p = 0.063$) and spatial ability ($F < 1$). Thus, there was also no need to include those variables as covariates into the calculations. However, correlations between the main outcome variables and the covariates are presented in Table 2.

2.2.2. Learning outcomes

Learning outcomes were analysed by two-factorial ANOVAs with the independent variables visualization (no visualization vs. static pictures vs. animation) and narration (no narration vs. non-process narration vs. process narration).

With respect to the Introduction-Score that referred to content that was not directly subject to experimental manipulation, a 3x3-ANOVA revealed neither a main effect for narration ($F < 1$) nor an interaction between visualization and narration ($F(4, 239) = 1.78, p = 0.13$), but a main effect for visualization ($F(2, 239) = 3.72, p = 0.026, \eta^2_p = 0.030$). This surprising effect, as all groups received identical information in the same format, was traced back with Bonferroni-adjusted simple comparisons. These results revealed that learners with animations performed better on the Introduction-Score than learners with no visualizations ($p = 0.002$) or static pictures ($p = 0.016$), whereas all other comparisons were not statistically significant.

With respect to the Comprehension-Score that addressed content that was subject to experimental manipulation, a 3×3 -ANOVA with the independent variables of visualization (no visualization vs. static pictures vs. animation) and narration (no narration vs. non-process narration vs. process narration) was conducted (see Fig. 2). Results revealed a strong main effect for visualizations ($F(2, 243) = 54.14, p < 0.001, \eta^2_p = 0.308$), a main effect for narration ($F(2, 243) = 7.87, p < 0.001, \eta^2_p = 0.061$), but no interaction between visualization and narration ($F < 1$).

The main effect of visualizations for the Comprehension-score was analysed by means of planned contrasts that corresponded to Hypothesis 1a (i.e., the text-only conditions compared to the visualization-plus-narration conditions) and Hypothesis 2a (i.e., the static picture conditions compared to the animation conditions), respectively. In line with Hypothesis 1a, planned contrasts showed a significant effect ($t(243) = 8.32, p < 0.001, \eta^2_p = 0.335$) indicating that learners in the visualization-plus-narration conditions outperformed learners in the text-only conditions. In line with Hypothesis 2a, planned contrasts showed that learners in the animation conditions outperformed learners in the static picture conditions ($t(243) = 2.47, p = 0.014, \eta^2_p = 0.033$).

Similarly, the main effect of narration was analysed by means of planned contrasts. In line with Hypothesis 1b, planned contrasts showed a significant effect ($t(243) = 2.88, p = 0.004, \eta^2_p = 0.048$) indicating that learners in the conditions with narrations outperformed learners that did not receive narrations. Moreover, there was a significant difference between the two narrated versions, with learners receiving process descriptions in the narration outperforming learners receiving non-process descriptions ($t(243) = 2.12, p = 0.035, \eta^2_p = 0.027$).

2.2.3. Cognitive load

Perceived difficulty and mental effort as components of cognitive load (during learning and during testing, respectively) were analysed by two-factorial ANOVAs with the dependent variable Comprehension-score and the independent variables visualization (no visualization vs. static pictures vs. animation) and narration (no narration vs. non-process narration vs. process narration).

Table 2
Correlations between main outcome variables and covariates (Experiment 1).

	Spatial ability	Intro-duction score	Compre-hension score	Difficulty learning	Difficulty testing	Mental effort learning	Mental effort testing
Prior knowledge	r 0.05	0.19**	0.17**	-0.15*	-0.16*	0.00	0.07
Spatial Ability	r	0.27**	0.30**	-0.11	-0.16*	-0.10	0.00
Introduction Score	r		0.44**	-0.16*	-0.29**	0.06	0.14*
Comprehension Score	r			-0.28**	-0.49**	-0.06	-0.08
Difficulty learning	r				0.57**	0.11	0.12
Difficulty testing	r					0.21**	0.24**
Mental effort learning	r						0.46**

Note. * $p < 0.05$, ** $p < 0.01$.

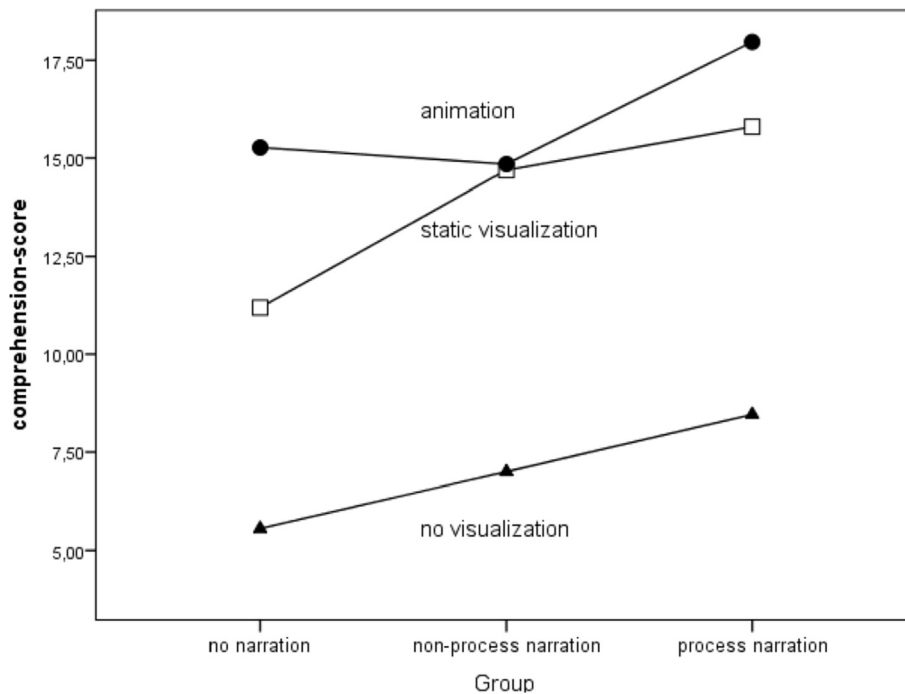


Fig. 2. Interaction graph with nine learning conditions and comprehension-score as dependent variable (Experiment 1).

With respect to perceived difficulty during learning, a 3×3 -ANOVA revealed a main effect for visualization ($F(2, 241) = 3.11, p < 0.046, \eta^2_p = 0.025$), no main effect for narration ($F < 1$) and no interaction between visualization and narration ($F < 1$).

With respect to perceived difficulty during testing, a 3×3 -ANCOVA revealed a main effect for visualization ($F(2, 243) = 9.61, p < 0.001, \eta^2_p = 0.073$), no main effect for narration ($F(2, 243) = 2.12, p = 0.123$) and no interaction between visualization and narration ($F < 1$).

The main effect of visualizations was analysed by means of planned contrasts that corresponded to Hypothesis 1c (i.e., the text-only conditions compared to the visualization-plus-narration conditions) and Hypothesis 2b (i.e., the static picture conditions compared to the animation conditions), respectively. In line with Hypothesis 2b, planned contrasts showed that learners in the animation conditions perceived less difficulty during learning ($t(241) = 2.30, p = 0.023, \eta^2_p = 0.030$) and during testing ($t(243) = 2.96, p = 0.003, \eta^2_p = 0.048$) than learners in the static-picture conditions.

With respect to Hypothesis 1c, planned contrasts showed no significant effect ($t(241) = 0.54, p = 0.590$) during learning but during testing ($t(243) = 2.16, p = 0.032, d = 0.37$) indicating that learners in the visualization-plus-narration conditions did not perceive less difficulty during learning but less difficulty afterwards during testing.

With respect to mental effort, a 3×3 -ANOVA with the independent variables visualization (no visualization vs. static pictures vs. animation) and narration (no narration vs. non-process narration vs. process narration) was conducted. Results revealed no main effect for visualization ($F < 1$), no main effect for narration ($F < 1$) and no interaction between visualization and narration ($F < 1$).

The Introduction-score was negatively related to perceived difficulty during learning ($r = -0.16, p < 0.05$) and to perceived difficulty during testing ($r = -0.29, p < 0.001$). The Comprehension-score was also negatively related to perceived difficulty during learning ($r = -0.28, p < 0.001$) and perceived difficulty during testing ($r = -0.49, p < 0.001$), meaning the more difficult learners found it to learn with the material or fill in the post knowledge test, the lower were their learning outcomes. Mental effort during learning was neither related to the Comprehension-Score nor to the Introduction-Score ($ps > 0.05$). Mental effort during testing was not related to Comprehension-score ($p > 0.05$), but weakly related to the Introduction-score ($r = 0.14, p < 0.05$), meaning that higher mental effort resulted in higher learning outcomes in the Introduction-score.

Perceived difficulty during learning did not correlate with mental effort during learning and mental effort during testing ($ps > 0.05$). However, there were significant correlations between perceived difficulty during testing and mental effort during learning ($r = 0.21, p < 0.01$) and mental effort during testing ($r = 0.24, p < 0.001$), meaning that participants invested more mental effort in case of higher perceived difficulty in the testing phase.

2.3. Discussion of experiment 1

In Experiment 1, the interplay of the information in visualizations (no visualization vs. static pictures vs. animation) and the information in narrations (no narration vs. non-process narration vs. process narration) was investigated. Firstly, in line with the multimedia principle and in line with theoretical expectations (Hypothesis 1a and 1b), it was shown that the presence of visualizations as well as the presence of narrations was necessary to achieve a better understanding of the content. As the two knowledge tests revealed, this was specifically the case for content that was subject to the experimental manipulation. Secondly, there was an instructional advantage of animations over static pictures (Hypothesis 2a). Furthermore, process narrations were significantly superior to no narration and to non-process narrations conditions. However, thirdly, the hypothesized interaction between type of narration and type of visualizations was not observed: Contrary to our expectations, the instructional advantage of animations over static pictures was not more pronounced for non-process descriptions in the narration compared to process descriptions.

Somewhat surprisingly, the contents of the introductory material, which were not subject to experimental variation and identical for all participants, were poorly understood by participants learning with static pictures or without visualizations in the experimental phase. It might be the case that these learners, who were disadvantaged in the experimental phase and thus displayed a lack of understanding for the issues to be learnt, then had trouble to fully understand what they had learnt before – whereas the others were at an advantage due to their superior learning advantages in the second phase, which allowed them to interconnect their knowledge of both phases facilitating retention and comprehension of the issues to be learnt. This is not implausible since learners were free to navigate within the computerized instructional material and to go back to the introductory material after they had viewed the experimentally varied content. Furthermore, successful learning during the experimental phase might have had an effect on the learners' motivation and might also have encouraged them to go back once more to the first phase in order to revise what they had done there or clarify aspects that were still unclear to them.

With regard to the subjective measure of perceived difficulty, the results were partly in line with the results of the learning outcomes: Participants in the animations conditions reported less difficulty during the learning and the testing phase than participants in the static pictures conditions. The multimedia-effect was, however, only confirmed by the difficulty during testing measure. Participants without visualizations found the testing phase more difficult than participants who had learnt with visualizations. However, perceived difficulty did not reflect the influence of the narration. The subjective measure of mental effort was unrelated to the comprehension of the content. Thus, the amount of process information presented by the narration did not affect the superiority of animation, neither when considering learning outcome nor cognitive load during learning or during testing.

As stated before, we had to exclude participants with too much prior knowledge in order to be able to compare participants with equal prerequisites across the different conditions. Although the sample size was big enough, this unfortunate concurrence reduced both the validity and the generalizability of our results. Therefore, we decided to conduct a replication of Experiment 1. However, for economic reasons, we refrained from investigating conditions without visualizations again, since we observed a huge effect size concerning this issue and since the importance of the multimedia principle has been shown in many studies before. Furthermore, the condition no visualization/no narration was ethically doubtful. Nevertheless, we still kept the conditions without narrations, since the effects were not as strong – particularly with respect to the comparison of non-process narration to process narration. Also, the investigation of this specific type of multimedia principle is under-represented in research.

3. Experiment 2

3.1. Method

3.1.1. Participants and design

We calculated the required sample size for the 2x3-Between Subject Design with the software tool G*Power 3.1 (Faul et al., 2009). Assuming a medium effect size $f = 0.25$, a Type I error of 0.05 and a Type II error of 0.20, the minimum required total sample size was 158 participants. Again more participants were used.

Finally, 181 seventh and eighth graders from two German secondary schools (54.4% female; $M = 13.08$ years, $SD = 0.71$, eleven different classes) participated in the study and were randomly assigned to one of six conditions, which resulted from a 2x3-design with visualization (static pictures vs. animation) and narration (no narration vs. non-process narration vs. process narration) as independent variables. The numbers of participants for each condition are presented in Table 3.

As stated above, learning about surfactants is content of ninth grader's curriculum in Germany. Therefore, seventh and eighth graders were again asked to take part in our study to ensure participants' low prior knowledge. The students indeed only achieved on average $M = 1.35$ ($SD = 1.06$) points in the prior knowledge test, whereas 18 points were attainable.

3.1.2. Instructional materials

The same computerized instructional material as in Experiment 1 was used, with the exception that there were no conditions without visualizations (i.e., three conditions less than in Exp. 1).

Table 3
Means (and SD) as a function of text version and visualization format (Experiment 2).

Text version	No narration		Non-process description		Process description	
Type of visualization	Static pictures	Animation	Static pictures	Animation	Static pictures	Animation
<i>n</i>	32	30	33	29	28	29
Control variables						
Prior knowledge	1.23 (0.94)	1.67 (1.12)	1.11 (1.09)	1.42 (1.03)	1.19 (1.13)	1.43 (1.03)
Spatial abilities	4.19 (1.82)	4.33 (1.42)	4.49 (1.97)	4.38 (1.80)	4.68 (1.39)	4.69 (1.54)
Cognitive Load						
Difficulty learning	5.22 (2.00)	5.20 (1.73)	5.72 (1.94)	5.83 (1.67)	5.75 (1.88)	5.10 (1.52)
Difficulty testing	7.42 (1.39)	7.21 (1.05)	7.46 (1.20)	7.04 (1.50)	7.46 (1.37)	7.04 (1.43)
Effort learning	5.56 (2.11)	5.40 (1.45)	5.73 (1.91)	6.28 (1.36)	5.75 (1.90)	5.93 (1.91)
Effort testing	7.28 (1.30)	6.60 (1.79)	6.42 (1.70)	6.10 (1.95)	6.11 (2.03)	6.44 (2.04)
Learning outcomes						
Introduction	2.88 (2.28)	2.83 (2.65)	2.21 (2.29)	2.45 (2.20)	2.82 (2.20)	3.17 (2.54)
Comprehension	11.13 (4.23)	13.70 (5.11)	12.30 (5.33)	14.55 (6.23)	13.36 (7.18)	15.45 (4.72)

3.1.3. Measures

The same paper-based measures as in Experiment 1 were used (prior knowledge test, spatial ability test, and subjective ratings of cognitive load).

3.1.4. Procedure

The same procedure as in Experiment 1 was used with the exception that cognitive load during testing was measured only once at the end. Similarly to Experiment 1, four different constructs are reported: the load (perceived difficulty and mental effort, respectively) experienced during learning and the load (again, perceived difficulty and mental effort, respectively) experienced during testing.

3.1.5. Statistical methods

Data was analysed within the framework of the General Linear Model (Horton, 1978) using analyses of variance (ANOVA). A calculation of intraclass correlation coefficients (ICC) again revealed that the data was not hierarchically clustered regarding the different school classes participants were from (ICCs < 0.001). Thus, multi-level analyses were not necessary. Regarding the main outcome variable, we also checked for assumptions of ANOVA— homogeneity of variance, normality, and independence of observations. In three subgroups, the assumption of normality was violated (measured with a Kolmogorov-Smirnov test). However, ANOVAs are considered as fairly robust against violations of normality (Glass et al., 1972).

3.2. Results of experiment 2

Means and standard deviations are reported in Table 3. For all statistical tests a significance level of 0.05 was applied. For ANOVAs and *t*-tests, partial eta-squared (η^2_p) is reported as a measure of effect size.

3.2.1. Control variables

With regard to the control variables, one-factorial ANOVAs revealed that learners' prior knowledge ($F(5, 161) = 1.06$, $p = 0.386$) and spatial abilities ($F < 1$) could be considered equal across the six experimental conditions. Therefore, no covariates were included in the analysis of the dependent variables learning outcomes and cognitive load. However, correlations between the main outcome variables and the background information are presented in Table 4.

3.2.2. Learning outcomes

Learning outcomes were analysed by two-factorial ANOVAs with the independent variables visualization (static pictures vs. animation) and narration (no narration vs. non-process narration vs. process narration).

With respect to the Introduction-Score that addressed content that was not subject to experimental manipulation, a 2×3 -ANOVA revealed no main effect for visualization ($F < 1$), no main effect for narration ($F(2, 175) = 1.33$, $p = 0.27$, $\eta^2_p = 0.015$) and no interaction between visualization and narration ($F < 1$).

With respect to the Comprehension-Score that addressed content that was subject to experimental manipulation, a 2×3 -ANOVA revealed, like in Experiment 1, a main effect for visualization ($F(1, 175) = 7.88$, $p = 0.006$, $\eta^2_p = 0.043$) indicating that

Table 4
Correlations between main outcome variables and covariates (Experiment 2).

	Spatial ability	Intro-duction score	Compre-hension score	Difficulty learning	Difficulty testing	Mental effort learning	Mental effort testing
Prior knowledge	<i>r</i> 0.01	0.04	0.13	0.00	-0.10	0.11	0.00
Spatial Ability	<i>r</i>	0.07	0.28**	0.15*	0.08	0.18*	0.12
Introduction Score	<i>r</i>		0.32**	-0.10	-0.20*	-0.06	0.07
Comprehension Score	<i>r</i>			-0.05	-0.31**	0.08	0.15*
Difficulty learning	<i>r</i>				0.39**	0.52**	-0.03
Difficulty testing	<i>r</i>					0.26**	0.18*
Mental effort learning	<i>r</i>						0.29**

Note. * $p < 0.05$, ** $p < 0.01$.

animations were superior to static pictures (Hypothesis 2a). There was no main effect for narration ($F(2, 175) = 1.94$, $p = 0.15$, $\eta^2_p = 0.022$) and no interaction between visualization and narration ($F < 1$). This indicates, however against our assumptions (Hypothesis 3), that the amount of process information presented by the narration did not affect learning with animation versus static pictures (see Fig. 3).

The influence of narration was analysed by means of planned contrasts (even though the main effect did not reach significance). In line with Hypothesis 1b, planned contrasts showed a significant effect ($t(175) = 1.74$, $p = 0.042$ (one-tailed), $\eta^2_p = 0.018$) indicating that learners in the conditions with narrations outperformed learners that did not receive narrations. However, learners receiving process descriptions in the narration did not outperform learners receiving non-process descriptions ($F < 1$).

3.2.3. Cognitive load

Perceived difficulty during learning and during testing as well as mental effort during learning and during testing were analysed by two-factorial ANOVAs with the dependent variable Comprehension-score and with the independent variables visualization (static pictures vs. animation) and narration (no narration vs. non-process narration vs. process narration).

With respect to perceived difficulty during learning, results revealed no main effect for visualization ($F < 1$), no main effect for narration ($F(2, 174) = 1.53$, $p = 0.22$) and no interaction between visualization and narration ($F < 1$).

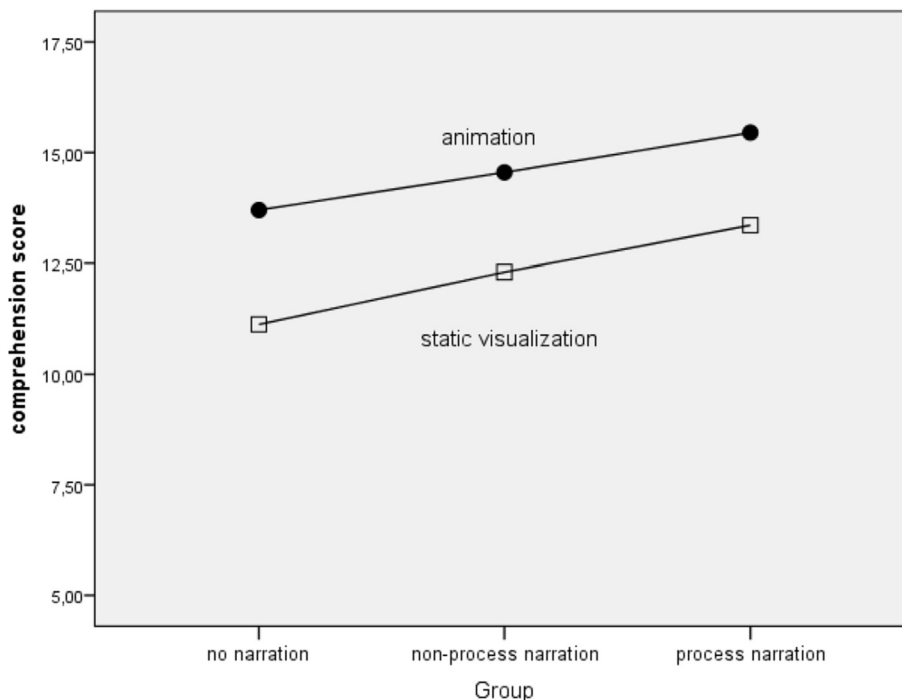


Fig. 3. Interaction graph with six learning conditions and comprehension-score as dependent variable (Experiment 2).

With respect to perceived difficulty during testing, results revealed only a marginally significant main effect for visualization ($F(1, 172) = 3.12, p = 0.08, \eta^2_p = 0.02$), but no main effect for narration ($F < 1$) and no interaction ($F < 1$).

With respect to mental effort, a 2x3-ANOVA with the dependent variable Comprehension-score and the independent variables visualization (static pictures vs. animation) and narration (no narration vs. non-process narration vs. process narration) was conducted. For mental effort during learning, results revealed no main effect for visualization, no main effect for narration and no interaction between visualization and narration (all $F_s < 1$).

With respect to mental effort during testing, results revealed no main effect for visualization ($F < 1$) and a marginally significant main effect for narration ($F(2, 172) = 2.79, p = 0.06, \eta^2_p = 0.03$). Helmert contrasts indicate that participants in the no narration condition invested more mental effort than participants in the static pictures and the animation conditions ($p = 0.02$). However, there was no interaction between visualization and narration ($F(2,172) = 1.16, p = 0.32, \eta^2_p = 0.01$).

The Introduction-score was not related to perceived difficulty during learning, but negatively related to perceived difficulty during testing ($r = -0.21, p < 0.01$). The Comprehension-score was also not related to perceived difficulty during learning, but negatively related to perceived difficulty during testing ($r = -0.31, p < 0.001$), meaning the more difficult learners found it to fill in the post knowledge test, the lower were their learning outcomes. Mental effort during learning was neither related to the Comprehension-Score nor to the Introduction-Score ($ps > 0.05$). Mental effort during testing was not related to Introduction-score ($p > 0.05$), but loosely related to the Comprehension-score ($r = 0.15, p < 0.05$), meaning that higher mental effort resulted in higher learning outcomes in the Comprehension-score.

Perceived difficulty during learning correlated with mental effort during learning ($r = 0.52, p < 0.001$), but not with mental effort during testing ($p > 0.05$). Furthermore, there were significant correlations between perceived difficulty during testing and mental effort during learning ($r = 0.26, p < 0.001$) and mental effort during testing ($r = 0.18, p < 0.05$), meaning that participants invested more mental effort in case of higher perceived difficulty in the testing phase.

3.3. Discussion of experiment 2

The findings of Experiment 1 could only partly be replicated in Experiment 2. Firstly, regarding Hypothesis 1b and 1c, receiving (any) narrations was better than receiving no narrations, which provides further empirical evidence for this specific case of the multimedia principle (even though the main effect was not significant). This finding was strengthened by the participants' mental effort during testing: Participants learning with no narration invested more mental effort during testing than participants learning with (any) narration.

Secondly, in line with Hypothesis 2a and as in Experiment 1, learners receiving animations outperformed learners receiving static pictures in the comprehension test.

Thirdly, other than expected (Hypothesis 3), but in line with the results of Experiment 1, there was no interaction observable between visualization and narration. Concerning the comparison of process descriptions and non-process descriptions in narrations, no significant differences were observable, in contrast to Experiment 1. As in Experiment 1, the subjective measure of perceived difficulty was negatively related to measures of learning outcome. However, contrary to Experiment 1, the perceived difficulty measure did not reflect the results of the comprehension test.

4. Limitations

There are a few methodological aspects limiting the generalizability of the underlying study. First, in both experiments, the assumptions of the ANOVAs were not totally fulfilled. Nevertheless, due to the facts that ANOVAs are considered as fairly robust against violations of normality (Glass et al., 1972) and that there are no non-parametrical methods examining two factor analyses, we felt justified in conducting ANOVAs.

Second, in the first Experiment, the randomization process was not successful resulting in covariates that were not equally distributed over the nine learning conditions. We had to exclude 31 participants which led to a lower generalizability of our results. Therefore, further experiments examining this research question might be useful.

Third, we used single items for measuring cognitive load based on the original Paas (1992) scales. More recent research suggests that multi-item scales are more sensitive to individual cognitive loads and reliable (Leppink, Paas, van Gog, van der Vleuten, & van Merriënboer, 2014). For future research, such scales measuring different loads should be used.

Finally, in our two experiments, participants had to learn in a self-regulated manner. The time limit was ten minutes. However, some learners could have used less time to learn, which may have influenced final learning outcomes. As time on task was not measured, we do not know if learning time varied or made an impact. Therefore, future studies examining the effects of narrations on learning with animations compared to static pictures should take into account how long participants are actively processing the information.

5. General discussion

Empirical results on learning with animations compared to static pictures are somewhat mixed. However, a meta-analysis by Höffler and Leutner (2007) revealed a medium advantage of animations over static pictures. This superiority was especially the case when processes were to be learned. In this context, one could assume that annotating narrations plays an important role as they provide further information about the process in question, which also might be important to successfully

understand the visualized process. Due to a lack of research on this topic, the aim was to investigate the role of process information in annotating narrations when learning with animations and static pictures. The underlying question was whether the amount of process information in narrations might affect learning with animations compared to static pictures differently.

In two experiments, seventh and eighth graders had to learn about chemical processes during washing laundry and were randomly assigned to different learning conditions which resulted from the combination of different types of visualizations (study 1: no visualization vs. static pictures vs. animation; study 2: static pictures vs. animation) with different types of narrations (no narration vs. non-process narration vs. process narration).

As the data revealed, not all of the theoretical expectations could be confirmed. Firstly, a necessary precondition for analyzing more subtle effects of visualizations and their moderators is that visualizations added to text should enhance learning. In line with our expectations which were based on the multimedia principle (Mayer, 2009), it could be shown that the presence of visualizations as well as the presence of narrations facilitated a better understanding of the content. As the different knowledge tests showed, this was specifically the case for content that was subject to experimental manipulation.

Somewhat surprisingly, the contents of the introductory material, which were not subject to the experimental variation and identical for all participants, were understood worse by participants learning with static pictures or without visualizations in the experimental phase. Due to the fact that those results were only significant in Experiment 1, this might be considered an empirical artefact. This once again stresses the importance of replicating experiments – as done in the current project – as well as the need for further research.

Secondly, the animations from the instructional material we used had already been proven to possess instructional advantages over static pictures (e.g. Höffler & Leutner, 2011), so we expected to be able to replicate these findings. Both experiments revealed exactly this advantage.

Thirdly, it was hypothesized that the kind of textual information of an accompanying narration would moderate learning with animations and static pictures. In line with the redundancy principle (Kalyuga & Sweller, 2014), it was expected that the instructional superiority of animations over static pictures might be even more pronounced when the accompanying text does not contain redundant process information. The hypothesized interaction between the type of narration and the type of visualizations was not observable in our data, however. One reason could be that the process was easy to understand, so that the visualization was sufficient and the narration was redundant (or the other way around). However, our data does not confirm this explanation attempt. Participants rated perceived difficulty during learning and testing with a mean around 6 (on a scale between 1 and 9), implying a rather high difficulty.

Alternatively, it may mean that the text could simply not be regarded as redundant, and that the process information contained in the narration was helpful in understanding the chemical reactions taking place for these students. One might even argue that our data gives some indication that process narration is not redundant at all but even generally helpful for both static pictures and animations (significant main effect in Experiment 1). For future research, more tailor-made instructional materials should be used with a more thorough investigation into potential redundancy effects in order to examine these effects.

In terms of cognitive load, results from Experiment 1 show that perceived difficulty ratings were partly in line with learning outcome measures and therefore support the basic findings. Yet, mental effort did not vary with the experimental manipulation. In this context, the question of how and when to measure cognitive load is still a topic of major interest (cf. de Jong, 2010). In both experiments, perceived difficulty was consistently positively correlated with mental effort and negatively correlated with learning outcome.

In terms of learning with animations and static pictures, a further moderator should be taken into account in future research: motivation. It might be expected that a more cognitive-motivational than a simple cognitive view on learning processes could enrich empirical research on learning with animations and static pictures. The appealing character of animations could, for instance, function as a compensator that overshadows higher demands in comparison to static pictures.

Although the expected moderating effect of narrations on learning with animations and static pictures could not be confirmed, the idea should not be rejected. Instead, future research should incorporate studies in which the role of process information in annotating narrations or written texts when learning with animations and static pictures is examined. The effects may be very subtle and sensitive to various moderating factors. Therefore, the current study could be seen as a starting point to generate a more sophisticated view on the interaction between visualization and narrations in learning processes – or, possibly, on a general positive influence of additional process information for visualizations.

Acknowledgments

This research was partially supported by a Center of Educational Studies scholarship from Ruhr-University Bochum.

Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.compedu.2016.11.001>.

References

- Adesope, O., & Nesbit, J. C. (2012). Verbal redundancy in multimedia learning environments. A meta-analysis. *Journal of Educational Psychology*, 104, 250–263.
- Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction*, 16, 183–198.
- Ayres, P., & Paas, F. (2007). Can the cognitive load approach make instructional animations more effective? *Applied Cognitive Psychology*, 21, 811–820.
- Baddeley, A. (1986). *Working memory*. Oxford: Oxford University Press.
- Berney, S., & Bétrancourt, M. (2016). Does animation enhance learning? A meta-analysis. *Computers & Education*, 101, 150–167.
- Bétrancourt, M. (2005). The animation and interactivity principles in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 287–296). Cambridge: Cambridge University Press.
- Boucheix, J. M. (2008). Young learner's control of technical animations. In R. Lowe, & W. Schnotz (Eds.), *Learning with animation* (pp. 208–234). Cambridge: Cambridge University Press.
- Castro-Alonso, J. C., Ayres, P., & Paas, F. (2014). Learning from observing hands in static and animated versions of non-manipulative tasks. *Learning and Instruction*, 34, 11–21.
- Catrambone, R., & Seay, A. (2002). Using animation to help students learn computer algorithms. *Human Factors*, 44, 495–511.
- Chandler, P. (2004). The crucial role of cognitive processes in the design of dynamic visualizations. *Learning and Instruction*, 14, 353–357.
- Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, 8, 293–332.
- Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Manual for kit of factor-referenced cognitive tests*. Princeton: Educational Testing Service.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41, 1149–1160.
- Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Review of Educational Research*, 42, 237–288.
- Hegarty, M. (1992). Mental animation: Inferring motion from static displays of mechanical systems. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 1084–1102.
- Hegarty, M., Kriz, S., & Cate, C. (2003). The roles of mental animations and external animations in understanding mechanical systems. *Cognition and Instruction*, 21, 325–360.
- Höffler, T. N. (2010). Spatial ability: Its influence on Learning with visualizations - a meta-analytic review. *Educational Psychology Review*, 22, 245–269.
- Höffler, T. N., & Leutner, D. (2007). Instructional animation versus static pictures: A meta-analysis. *Learning and Instruction*, 17, 722–738.
- Höffler, T. N., & Leutner, D. (2011). The role of spatial ability in learning from instructional animations: Evidence for an ability-as-compensator hypothesis. *Computers in Human Behavior*, 27, 209–216.
- Höffler, T. N., Schmeck, A., & Opfermann, M. (2013). Static and dynamic visual representations: Individual differences in processing. In G. Schraw, M. T. McCrudden, & D. Robinson (Eds.), *Learning through visual displays* (pp. 133–163). Charlotte, NC: Information Age Publishing.
- Höffler, T. N., & Schwartz, R. (2011). Effects of pacing and cognitive style across dynamic and non-dynamic representations. *Computers & Education*, 57, 1716–1726.
- Horton, R. L. (1978). *The general linear model: Data analysis in the social and behavioral sciences*. London: McGraw-Hill.
- de Jong, T. (2010). Cognitive load theory, educational research, and instructional design: Some food for thought. *Instructional Science*, 38, 105–134.
- Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology*, 13, 351–371.
- Kalyuga, S., Chandler, P., & Sweller, J. (2000). Incorporating learner experience into the design of multimedia instruction. *Journal of Educational Psychology*, 92, 126–136.
- Kalyuga, S., & Sweller, J. (2014). The redundancy principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 247–262). New York: Cambridge University Press.
- Kühl, T., Scheiter, K., Gerjets, P., & Edelman, J. (2011). The influence of text modality on learning with static and dynamic visualizations. *Computers in Human Behavior*, 27, 29–35.
- Large, A., Beheshti, J., Breuleux, A., & Renaud, A. (1996). Effect of animation in enhancing descriptive and procedural texts in a multimedia learning environment. *Journal of the American Society for Information Science*, 47, 437–448.
- Lee, D. Y., & Shin, D. H. (2011). Effects of spatial ability and richness of motion cue on learning in mechanically complex domain. *Computers in Human Behavior*, 27, 1665–1674.
- Leppink, J., Paas, F., van Gog, T., van der Vleuten, C. P. M., & van Merriënboer, J. J. G. (2014). Effects of pairs of problems and examples on task performance and different types of cognitive load. *Learning and Instruction*, 30, 32–43.
- Lowe, R. K. (2008). Learning with animation. Where to look, when to look. In R. Lowe, & W. Schnotz (Eds.), *Learning with animation* (pp. S. 49–70). Cambridge: Cambridge University Press.
- Lowe, R. K., & Schnotz, W. (2014). Animation principles in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 513–546). New York: Cambridge University Press.
- Low, R., & Sweller, J. (2014). The modality principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 227–246). New York: Cambridge University Press.
- Mayer, R. E. (1989). Systematic thinking fostered by illustrations in scientific text. *Journal of Educational Psychology*, 81, 240–246.
- Mayer, R. E. (2001). *Multimedia learning*. Cambridge: Cambridge University Press.
- Mayer, R. E. (2006). Ten research-based principles of multimedia learning. In H. F. O'Neil (Hrsg.) (Ed.), *Web-based learning* (pp. 371–390). Mahwah, NJ: Erlbaum.
- Mayer, R. E. (2009). *Multimedia learning*. Cambridge: Cambridge University Press.
- Mayer, R. E., & Anderson, R. B. (1991). Animations need narrations: An experimental test of a dual-coding hypothesis. *Journal of Educational Psychology*, 83, 484–490.
- Mayer, R. E., & Anderson, R. B. (1992). The instructive animation: Helping students build connections between words and pictures in multimedia learning. *Journal of Educational Psychology*, 84, 444–452.
- Mayer, R. E., Hegarty, M., Mayer, S., & Campbell, J. (2005). When static media promote active learning: Annotated illustrations versus narrated animations in multimedia instruction. *Journal of Experimental Psychology: Human Perception and Performance*, 11, 256–265.
- Moreno, R., & Mayer, R. E. (2002). Learning science in virtual reality multimedia environments: Role of methods and media. *Journal of Educational Psychology*, 94, 598–610.
- Narayanan, N. H., & Hegarty, M. (1998). On designing comprehensible interactive hypermedia manuals. *International Journal of Human-Computer Studies*, 48, 267–301.
- Narayanan, N. H., & Hegarty, M. (2002). Multimedia design for communication of dynamic information. *Journal Human-Computer Studies*, 57, 279–315.
- Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, 84, 429–434.
- Paivio, A. (1986). *Mental representations. A dual coding approach*. Oxford: Oxford University Press.
- Plötzner, R., Bodemer, D., & Neudert, N. (2008). Successful and unsuccessful use of dynamic visualisations in instructional texts. In R. Lowe, & W. Schnotz (Eds.), *Learning with animation* (pp. 71–91). Cambridge: Cambridge University Press.
- Plötzner, R., & Lowe, R. (2004). Dynamic visualisations and learning. *Learning and Instruction*, 14, 235–240.
- Rieber, L. P. (1990). Animation in computer-based instruction. *Educational Technology Research and Development*, 38, 77–86.

- Rieber, L. P., & Kini, A. S. (1991). Theoretical foundations of instructional applications of computer-generated animated visuals. *Journal of Computer-Based Instruction*, 18, 83–88.
- Roscoe, R. D., Jacovina, M. E., Harry, D., Russell, D. G., & McNamara, D. S. (2015). Partial verbal redundancy in multimedia presentations for writing strategy instruction. *Applied Cognitive Psychology*, 29, 669–679.
- Schmidt-Weigand, F., & Scheiter, K. (2011). The role of spatial descriptions in learning from multimedia. *Computers in Human Behavior*, 27, 22–28.
- Schnotz, W., & Lowe, R. K. (2008). A unified view of learning from animated and static graphics. In R. Lowe, & W. Schnotz (Eds.), *Learning with animation* (pp. 304–356). Cambridge: Cambridge University Press.
- Schwan, S., & Riempp, R. (2004). The cognitive benefits of interactive videos: Learning to tie nautical knots. *Learning and Instruction*, 14, 293–305.
- Slavin, R. E. (1986). Best-evidence synthesis: An alternative to meta-analytic and traditional reviews. *Educational Researcher*, 15, 5–11.
- Sweller, J. (1988). Cognitive load during problem solving. Effects on learning. *Cognitive Science*, 12, 257–285.
- Sweller, J. (1999). *Instructional design in technical areas*. Melbourne: ACER Press.
- Sweller, J. (2010). Cognitive load theory. Recent theoretical advances. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory* (pp. S. 29–47). Cambridge: Cambridge University Press.
- Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York: Springer.
- Tversky, B., Morrison-Bauer, J. B., & Bétrancourt, M. (2002). Animation: Can it facilitate? *International Journal of Human-Computer Studies*, 57, 247–262.
- Wong, A., Leahy, W., Marcus, N., & Sweller, J. (2012). Cognitive load theory, the transient information effect and e-learning. *Learning and Instruction*, 22, 449–457.