

## Improved Effectiveness of Cueing by Self-Explanations when Learning from a Complex Animation

BJÖRN B. DE KONING<sup>1\*</sup>, HUIB K. TABBERS<sup>1</sup>, REMY M. J. P. RIKERS<sup>1</sup> and FRED PAAS<sup>1,2</sup>

<sup>1</sup>*Department of Psychology, Erasmus University Rotterdam, Rotterdam, The Netherlands*

<sup>2</sup>*Centre for Learning Sciences and Technologies, Open University of the Netherlands, Heerlen, The Netherlands*

*Summary: A major problem in learning from instructional animations is that the complex perceptual and cognitive processing exceeds the learner's limited processing capacities. Although attention cueing might help learners in focusing on essential parts of an animation, previous studies have shown that it does not necessarily improve learning performance. This study investigated whether generating self-explanations while studying a cued or an uncued animation might engage learners in cognitive activities necessary for learning. It was hypothesized that learning from a cued animation that reduces working memory load associated with searching for specific elements might be improved by generating self-explanations, whereas self-explaining with an uncued animation would have no positive effect on learning. The results confirmed the hypothesized interaction between cueing and self-explaining. They suggest that self-explanation enhances learning if visual cues are used to structure and highlight the essential parts of an animation. Copyright © 2010 John Wiley & Sons, Ltd.*

Instructional animations are increasingly used in computer-based learning environments because they extend the possibilities of static visualizations by explicitly depicting dynamic characteristics (i.e. motion, timing) of events. Intuitively, animations should therefore be well suited for enhancing learners' understanding of phenomena in which change-related information is essential for building an accurate mental representation. A recent meta-analysis by Höffler and Leutner (2007) supports this notion and provides evidence that under some conditions animations may be superior to static graphics.

However, most reviews have failed to establish clear advantages of animations over their static counterparts (e.g. Tversky, Morrison, & Bétrancourt, 2002). A number of reasons have been suggested for the ineffectiveness of animations. Ainsworth and VanLabeke (2004) have argued that because animations are transient, learners may miss or partially process essential information and therefore fail to construct an adequate mental representation. In addition, Lowe (1999) has suggested that animations increase demands on visual attention, as perceptually salient elements may attract learners' attention away from more subtle but thematically relevant elements of animations. Another suggestion is that animations seem to encourage passive rather than active learning (Hegarty, Kriz, & Cate, 2003). Whereas understanding the operation of complex dynamic systems from static graphics requires learners to infer the system's temporal changes through a process of internal animation, understanding the same system from animations requires learners to perceive the temporal changes by merely looking at an external display without necessarily having to engage in learning-related activities. Therefore, learners may be much less cognitively engaged in learning from animations than from statics. In the present study, we examined the instructional effectiveness of focusing

learners' attention on essential aspects of an animation when they are encouraged to actively process the presented information by generating self-explanations.

### Instructional animations and learning

Understanding complex dynamic systems, such as mechanical or biological systems, through animations requires learners to construct an internal mental representation that exactly represents the objects and their configuration, object behaviour (i.e. movements) and the causal relations in the systems' behaviour (Hegarty et al., 2003). A key aspect of the comprehension process involves accurately extracting the crucial information from the visual display, which may subsequently serve as input for further processing. However, there is currently no comprehensive theory that explains all perceptual and cognitive factors involved in learning from animations. Several theoretical models may be used. For example, Hegarty's (2005) general model of comprehension of visual displays explicitly identifies perceptual characteristics of elements in the visual display as important features that may affect what information is extracted and hence the quality of the mental representation that is constructed.

Alternatively, several more cognitively oriented theoretical models suggest that another aspect that may significantly affect the degree to which animations are processed satisfactorily is the learners' limited working memory capacity. According to cognitive load theory (CLT; Paas, Renkl, & Sweller, 2003) and Mayer's (2001) cognitive theory of multimedia learning animations may place high cognitive load on learners' cognitive resources associated with specific aspects of complex animations such as the temporally distributed nature and the high information load of the presentation. According to such an account, when animations fail to improve learning it is likely to be due in part to the high ineffective or extraneous load they produce (Ayres & Paas, 2007).

In a visual display, dynamic information such as movements and sudden appearances of objects effectively attracts

\*Correspondence to: Björn B. de Koning, Department of Psychology, Erasmus University Rotterdam, P.O. Box 1738, NL-3000 Rotterdam, The Netherlands. E-mail: dekonijng@fsw.eur.nl

human visual attention (Hillstrom & Chai, 2006). Therefore, animations showing many elements with different perceptual characteristics (i.e. colour, form, orientation), might contain several elements that simultaneously compete for attention, be it relevant for understanding the content or not. This may be especially overwhelming for novices as they do not know which elements should be attended to. For example, it has been found that brightly coloured elements that contrast with surrounding elements may easily misdirect learners' attention, forcing them to conduct searches to find relevant information (Lowe, 2003). The process of locating task-relevant information might be detrimental for learners' exploration of a complex animation because unnecessary visual searches require large amounts of working memory resources and therefore produce ineffective cognitive load (Ayres & Paas, 2007). Thus, processing animations with many (irrelevant) salient details might require so much cognitive resources that little remains for processing the actual subject-matter. Consequently, learners may fail to extract the information required for building a satisfactory mental representation of the animation. Furthermore, superficial processing of animations may occur because learners might study them in a passive viewing mode, that is, they perceive dynamic information from an external display without engaging in relevant cognitive activities (Hegarty *et al.*, 2003). This may mislead learners into believing that they have understood the content correctly (Rozenblit & Keil, 2002).

### Supporting learners' visual processing

Several researchers have argued that learning from animations may be enhanced when the visual search associated with splitting one's attention between and/or within representations is reduced by visual cues that direct learners' attention to an animation's essential parts (e.g. Bétrancourt, 2005). Particularly novices, who rely heavily on bottom-up processing and engage in high visual search at the cost of cognitive capacity, should benefit from cues for what information to select and how to organize it. By reducing visual search processes, ineffective mental activities that might overload working memory are reduced and cognitive resources for learning-related activities are freed. Thereby, learners' attention is directed to relevant parts, which provides an opportunity to concentrate on understanding an animation's content and underlying principles.

Using eye-tracking to examine attention allocation in learning from cued and uncued animations Kriz and Hegarty (2007) and De Koning, Tabbers, Rikers and Paas (in press) have demonstrated that attention-directing cues (i.e. arrow-cues and spotlight-cues) can effectively direct learners' attention towards specific elements in animations. It might be argued that focusing attention on different elements at different times may provide an organizing structure that allows learners to gradually build up an understanding of complex dynamic systems. However, cues emphasizing the elements that are crucial to understanding an animation do not ensure that conceptual relations between various elements are detected (De Koning *et al.*, in press; Kriz & Hegarty, 2007; but see De Koning, Tabbers, Rikers, & Paas,

2007). Thus, attention-directing cues seem *necessary* to help learners in extracting essential parts from animations, but cues are not *sufficient* to engage learners in the active knowledge construction necessary for learning.

### Increasing learners' cognitive engagement

Hence, for learning to occur in complex animations with cues, stimulating learners to engage in active processing will increase the probability that they understand the presented information (Mayer, 2001). Consequently, instructional techniques should be used that help learners to invest working memory resources in learning-related activities. A substantial number of studies have demonstrated that learners develop a deeper understanding of instructional materials if they generate explanations to themselves during learning (Aleven & Koedinger, 2002; Chi, de Leeuw, Chiu, & LaVancher, 1994; Renkl, Stark, Gruber, & Mandl, 1998). Several reasons have been proposed for why self-explaining improves learners' understanding. For example, self-explanations enable learners to integrate new knowledge with prior knowledge and/or with information within the learning material. Moreover, self-explaining forces learners to explicate their understanding, and to generate inferences, which might help them to monitor and revise their knowledge base (Renkl & Atkinson, 2002). These cognitive mechanisms may induce learners to construct a more integrated representation, yielding higher transfer of learning (Chi *et al.*, 1994). According to Renkl and Atkinson (2003), self-explanation activities may be considered effective or germane cognitive load because self-explaining increases cognitive load but directly contributes to mental model construction.

Recently, several studies have examined whether the format of instruction influences the effectiveness of self-explanations. Of most relevance to the present study are studies on self-explaining with diagrams. According to Cox (1999) diagrams facilitate self-explaining because compared to text, they require less mental computation and provide explicit feedback to which learners can compare their explanations. Comparing learners who self-explained with text or diagrams about the cardiovascular system, Ainsworth and Loizou (2003) provided evidence for this suggestion. However, Butcher (2006) argues that not all diagrams elicit meaningful self-explanations. She demonstrated that diagrams with attentional guidance to its essential parts and relations rather than diagrams in which these elements are less conspicuous are most effective for eliciting meaningful self-explanations and hence promoting understanding. A potential reason for this is that although generating self-explanations is a constructive activity, it requires considerable cognitive capacity (Renkl, 2005). This suggests that the effectiveness of self-explanations is influenced by the amount of learners' available working memory resources, and does not contribute to learning if total cognitive capacity is exceeded.

As the majority of studies have examined self-explaining in static visualizations and/or text, the present study is, to our knowledge, the first study investigating effects of self-explaining with an instructional animation. For learning

from animations, self-explanations are believed to improve conceptual understanding as it encourages learners to activate their knowledge of the depicted system and enables them to verify their knowledge against the actual processes explicitly displayed in the animation. Butcher's (2006) finding that attentional guidance to relevant information in a visualization enhances self-explaining suggests that self-explaining might especially be an effective cognitive activity to improve learning performances from cued animations. Cues that guide learners' attention to an animation's essential parts facilitate the selection and extraction of important elements thereby reducing extraneous cognitive load and, hence create a situation in which the engagement in learning-related activities (i.e. self-explaining) is more likely to occur effectively. In contrast, generating self-explanations while studying an uncued animation is less likely to improve learning performance due to the high extraneous load in the absence of attentional guidance. Thus, animations that reduce working memory demands associated with searching for relevant parts (i.e. extraneous load) and use techniques such as self-explanation for increasing cognitive engagement with genuine learning activities (i.e. germane load) may provide a highly effective instructional combination for improving learning (Paas, Renkl, et al., 2003).

In this study, learners studied a cued or an uncued animation about the cardiovascular system and did or did not generate self-explanations while viewing the animation. Because cues effectively capture attention (Kriz & Hegarty, 2007), spotlight-cues emphasizing the cardiovascular system's main subsystems were used in the cued conditions to focus learners' attention on relevant parts of the animation and to minimize processing of irrelevant aspects. Consequently, the high visual search and load associated with selecting essential information should therefore be reduced, thereby increasing working memory resources available for generating meaningful self-explanations (Renkl & Atkinson, 2003). Alternatively, it may be argued that consecutively cueing an animation's essential parts may provide a structured sequence of meaningful functional elements that learners can use to organize their self-explanation activities. This way, the amount of information that needs to be processed in working memory at one time is reduced, which provides an opportunity to generate meaningful self-explanations.

Because self-explaining is carried out in a continuous, ongoing and piecemeal fashion it gives rise to multiple opportunities to incrementally build up and revise one's internal representation that, ultimately, may lead to an integrated mental representation of the animation. Active processing of the presented information is thus crucial for learning, especially if the purpose is to generate deeper understanding of the material (Ainsworth & Loizou, 2003; Butcher, 2006). Therefore, it was expected that the effect of self-explaining with spotlight-cues would be reflected in all levels of understanding, but is likely to be more pronounced on measures of deep learning. Consequently, we hypothesized an interaction between cueing and self-explaining, indicating that learners studying a cued animation would benefit from self-explanations yielding higher performance on inference and transfer tests and to a lesser degree on a

retention test, whereas learners studying an uncued animation would not be able to profit from self-explanations and thus would not enhance their performance on these learning tests. In addition to the performance measures, we also examined the self-explanation protocols to see whether self-explaining with a cued or an uncued animation elicited qualitatively different self-explanations. It was hypothesized that spotlight-cues would promote the generation of meaningful self-explanations due to less visual search and increased working memory resources, indicating more meaningful self-explanations in the self-explanation protocols with a cued than with an uncued animation.

## METHOD

### Participants and design

The experiment conformed to a factorial design with the factors cueing (yes, no) and self-explaining (yes, no). Participants were 90 Dutch high school students (63 males and 27 females) with an age range between 13 and 15 years. Originally, 95 students participated, but due to data loss five students were eliminated from further analyses. Participants were randomly assigned to one of four conditions, which resulted in 21 participants in the condition with cueing and self-explaining, 25 in the condition with cueing and without self-explaining, 21 in the condition without cueing and with self-explaining, and 23 in the condition without cueing and without self-explaining. The experiment was part of an introductory biology course on the cardiovascular system. Although students had been introduced to the different components of the cardiovascular system, according to the teacher they were ignorant of the exact functioning of the system's dynamics (e.g. timing) at the time of the experiment. Participants had not been taught previous courses on the workings of the heart and the blood flow, so the subject-matter was new for them.

### Materials

The computer-based learning environment was developed with Macromedia Flash 7.0 (Macromedia, 2004). The materials were electronically presented on a 17" colour computer screen with a resolution of 1024 × 768 pixels. With the exception of the self-explanation data, all data were automatically recorded.

### *Demographic questionnaire*

The questionnaire, based on Mayer and Moreno (1998), asked participants to indicate their gender, age, years of high school education and experience with the cardiovascular system. Participants rated their experience with the cardiovascular system on a five-point scale, ranging from very little (1) to very much (5). The mean experience scores, i.e. 2.76 for the cued self-explaining condition, 2.32 for the uncued self-explaining condition, 2.35 for the no-explaining with cues condition and 2.14 for the no-explaining without cues condition did not differ,  $F(4, 97) = 1.14$ ,  $MSE = 0.85$ ,  $p > .05$ .

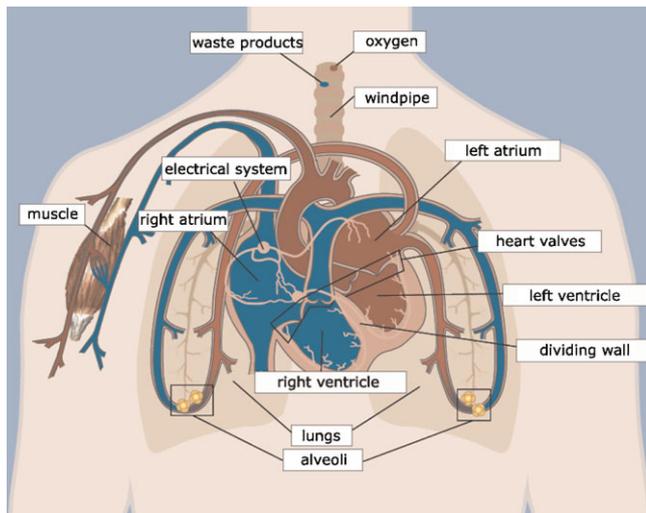


Figure 1. Labelled diagram of the cardiovascular system

### Static diagram

All participants studied a labelled static diagram which showed a coloured picture of the cardiovascular system in a relaxed state and included labels naming the system's parts (Figure 1). For the labels, general terms (e.g. valves) instead of specific terminology (e.g. tricuspid valve) were chosen to represent the structures of the cardiovascular system, because naming the system's structures in general terms avoids confusing different terms for the same concept. The labelled diagram ensured that all participants had learned the (same) prerequisite structural knowledge without which the self-explanation activity would be very difficult or perhaps be impossible. The static diagram was presented full-screen for 2 minutes. To assess whether participants were able to correctly recall the labels and match them to the appropriate structures, an unlabelled static diagram in which participants had to fill out the labels was used. To reduce the possibility that participants could rehearse the labels or name the structures on the basis of a sensory memory trace of the visual display, participants completed an intermediate task that required them to count backwards as from 585 in steps of three for 1 minute (Peterson & Peterson, 1959). An acceptance criterion of 95 per cent of correctly named structures was used for participants to continue with the experiment. If participants failed to meet this criterion, they were instructed to re-study the labelled diagram once and to fill out the unlabelled diagram again. Participants were discarded from the experiment if they did not meet the criterion after re-studying the materials. All participants did meet the criterion that was necessary to continue with the experiment.

### Practice exercise

The practice exercise consisted of studying an animation of how lightning develops with or without self-explaining for 2 minutes. Participants in the self-explaining conditions were given written instructions about self-explaining. Then they were instructed to self-explain during the animation in order to familiarize themselves with explaining aloud. In the no self-explaining conditions, participants only studied the animation without self-explaining. The practice exercise thus informed participants about the upcoming learning task (that

it closely resembled) and offered them the opportunity to practice and familiarize themselves with studying an animation with or without self-explaining.

### Animation

Two versions of an animation illustrating the workings of the human cardiovascular system were used. Both versions depicted the simultaneous dynamics of the cardiovascular system's main subsystems: Circulatory system, electrical system, pulmonary circulation, systemic circulation, valves system. Despite their individual properties and purposes, the subsystems are largely dependent on each other to let the cardiovascular system function correctly. The animation showed how the heart contracts, where blood flows, where oxygen and waste products are exchanged, and how the heart valves and electrical system work. The animation showed 26 cardiac cycles that each lasted approximately 12 seconds. It took 305 seconds to play the animation. The two animation versions were presented without accompanying written or verbal descriptions, contained no pauses, and learners could not control the animation. The animation's speed was set at 9.5 frames per second. Prior testing indicated that this speed enabled elaborated self-explanations, but was sufficiently challenging to prevent ceiling effects.

The only difference between the two animation versions was that in the cued version the different subsystems were sequentially cued, whereas the uncued version was presented without cues. Cueing was done by decreasing the luminance of all elements in the animation except the cued subsystem. This enables the cued subsystem to stand out against the rest of the animation and, therefore, to become more salient. However, all uncued elements were still visible and could be attended to. In the cueing conditions, each of the five subsystems was cued consecutively for 55 seconds with an inter-cueing period of 4.5 seconds in which the animation was not cued. Within each cueing period (i.e. 55 seconds), a single subsystem was cued and participants saw the cued subsystem in at least four cardiac cycles. Prior testing indicated that this cueing duration was necessary to be able to provide elaborated self-explanations on the cued subsystem. The first cue appeared 12 seconds after the animation started to ensure that participants would notice the visual contrast. Cueing followed a predetermined order (electrical system, circulatory system, pulmonary circulation, valves system, systemic circulation), which was derived from descriptions of the cardiovascular system in medical textbooks.

In the self-explanation conditions, participants explained aloud the functioning of the cardiovascular system while they viewed the animation. The cyclical nature of the animation allowed participants to self-explain without having to miss essential information. Self-explaining did not increase time on task, as participants only self-explained while they watched the animation, which had a fixed duration that was the same for all conditions. Furthermore, the experimenter did prompt participants to continue explaining when their comments were vague, incomplete, or when participants paused for more than 5 seconds. The prompts were based on the content-free prompts of Chi, Siler, Jeong, Yamauchi, and Hausmann (2001). Prompts did not give explanations, feedback or extra information to participants.

They were only intended to encourage participants to elaborate on the animation. Examples of prompts are, 'Could you clarify what you just said?' or 'Could you explain why it works like this?'. Prior to the experiment the experimenter was carefully instructed that he was not allowed to try to influence the learning process and should provide the prompts in a neutral way. The experimenter sat out of the participants' sight to minimize any uncomfortable feelings or pressure to perform well participants may have felt in the presence of the experimenter. All verbal comments were recorded with a tape recorder.

#### *Learning performance*

To assess participants' understanding of the cardiovascular system, 32 retention questions, five transfer questions (De Koning et al., in press), and 14 inference questions were used. An analysis of the animation in terms of the information that could be extracted from it ensured that all test items covered information that was presented in the animation or tested the knowledge that could be inferred from it.

The retention test consisted of multiple-choice items with four alternatives. Retention questions asked learners about information that was explicitly displayed in the animation. Comparable number of questions covered the structural and behavioural aspects of each subsystem. Test items asked about all relevant elements of a subsystem and because only complete subsystems were cued the questions directly corresponded to information in a cued region. To minimize the risk that questions were not fully understood, pictures were added that referred to elements mentioned in the question. Test items were only pictorial if questions were difficult to interpret with text alone or to clarify a concept. For instance, one retention question asked participants to choose the picture that correctly shows the direction of blood flow out of four pictures.

The inference test consisted of open-ended questions, which were specifically designed to assess the participants' ability to infer the animation's causal relations. Correctly answering the questions required participants to combine and integrate information from—different parts of—the animation and to relate this to learned information. For example, one question asked, 'What causes the valves of the heart to open?' Correctly answering this question required participants to recognize that closed valves prevent blood from flowing directly from the upper to the lower chambers, causing the upper chambers to build up pressure that, if high enough, opens the valves.

The transfer test consisted of open-ended questions that all covered the workings of the subsystems shown in the animation. Each question asked participants to reason about the functioning of one of the subsystems. These questions could only be answered correctly by applying the learned inferences about the cause-effect relations of the subsystems (i.e. cued parts) and, therefore, provides a good measure of the quality of the constructed representation (Mayer, 2001). An example of a transfer question is 'Explain why it is possible to divide the functioning of the electrical system into two steps.'

#### *Mental effort*

The mental effort measure used in this study was the nine-point subjective rating scale developed by Paas (1992),

ranging from very, very low mental effort (1) to very, very high mental effort (9). The scale is a reliable and valid measure of perceived working memory load (Paas, Tuovinen, Tabbers, & Van Gerven, 2003).

#### **Procedure**

The experiment was conducted in 10 sessions of 90 minutes. In each session between 1 and 15 participants were tested, with participants randomly assigned to one of the experimental conditions. Each participant was seated at a computer and completed the experiment at his or her own pace. Participants completed the experiment individually, but in some sessions participants in the no-explaining conditions completed the experiment simultaneously. Nevertheless, the experimenter was present and carefully monitored the progress of each of the participants in all experimental conditions. When participants had finished the experiment, they could do something by themselves in silence, but they were not allowed to leave the classroom or talk to other students.

First, participants filled out the demographic questionnaire. Subsequently, they studied the labelled diagram of the cardiovascular system followed by an unrelated number-counting task and completed a fill-in exercise to test whether they had correctly learned at least 95 per cent of the cardiovascular system's main structures. Participants then completed the practice exercise. Next, participants read instructions that stated that they should try to comprehend the content of the animation. Participants in the self-explanation conditions were also instructed to self-explain aloud while studying the animation. Depending on the experimental condition, one of the two animation versions was then presented on the computer. After the animation, participants indicated their invested mental effort by clicking on one of the options of the mental effort rating scale.

After an introduction to the test phase, the inference questions were presented that could be answered by typing the answer in an on-screen textbox. Inference items were given before retention items to ensure that answers to the inference questions could not be influenced by information provided in the multiple-choice retention questions. Subsequently, participants received the retention test. Retention questions could be answered by clicking on one of the four alternatives. Finally, participants completed the transfer questions, which they could answer using the keyboard. All test questions were individually and randomly presented on screen. Participants could sequentially move through the questions but they could not return to previously presented questions. Participants were allowed to take as much time as needed to answer each question. After each test, participants rated the invested mental effort.

#### **Analysis**

##### *Learning performance*

Participants' performance on retention, inference and transfer items was scored blind with respect to experimental condition. Each correctly answered retention item yielded one point, otherwise participants received zero points. The

maximum score was 26 points. The inference test was scored by the correct inferences for each question. For each correctly inferred relation on an inference question one point was awarded. Several questions required learners to make two inferences, yielding a maximum score of 23 points. Transfer questions were scored by counting the number of idea units per question produced by participants. For each question, one point was assigned to each correctly mentioned idea unit. The maximum score was 21 points.

#### Self-explanation protocols

Participants' think-aloud protocols were transcribed and coded according to a scheme based on Ainsworth and Burcham (2007). Modifications were made where necessary to adjust it to the present learning materials and the format of presentation. All verbal protocols were scored for statements corresponding to one of the following categories:

- 1) *Paraphrase*: Utterances were coded as paraphrases if participants merely verbalized information that was explicitly presented in the animation without adding new information in the form of an explanation. For example, the utterance 'blood flows from the atrium to the ventricle' would be coded within this category.
- 2) *Goal-driven explanation*: Explanations were considered goal-driven if participants made explanations that inferred a goal or function to a particular action or structure of the cardiovascular system. For example, saying that 'valves close to prevent blood from flowing upwards' would be attributing a purpose to the closing of the valves.
- 3) *Elaborative explanation*: If participants produced explanations that inferred information from the animation in an elaborated way, without assigning it a specific goal or purpose it was coded in this category. For example, the sentence 'blood in the muscle is oxygenated' would be placed in this category.
- 4) *Monitoring statements*: These statements indicated whether participants did or did not understand the information presented in the animation and reflected checks on understanding, confusion or questions about the learning

material. For example, utterances such as 'I do not understand how the electrical system works' or 'Okay, now I see why blood is traveling to the muscle' are assigned to this category.

- 5) *Inference errors*: If participants made a self-explanation that reflected incorrectly inferred knowledge, it was referred to as an incorrect self-explanation. Ambiguous or partially correct inferences were not counted as errors. For example, the sentence 'Blood entering the muscles takes up oxygen, and then flows back into the right side of the heart' was coded within this category because the purpose of the blood flow towards the muscles is inferred incorrectly.

## RESULTS

For all statistical tests, a significance level of .05 was applied. For any *post hoc* analyses, we used analysis of variance (ANOVA). Effect sizes are expressed in terms of partial eta squared ( $\eta_p^2$ ).

### Learning outcomes

Table 1 shows the mean scores and standard deviations on the dependent measures for all conditions. A MANOVA on the retention, inference and transfer scores with the between-participants factors cueing (yes, no) and self-explaining (yes, no) revealed a significant effect for self-explaining (Wilks'  $\lambda = 0.54$ ,  $F(3,84) = 24.11$ ,  $p < .05$ ,  $\eta_p^2 = .463$ ), and cueing (Wilks'  $\lambda = 0.90$ ,  $F(3,84) = 3.17$ ,  $p < .05$ ,  $\eta_p^2 = .102$ ) and a significant interaction-effect between cueing and self-explaining (Wilks'  $\lambda = 0.87$ ,  $F(3,84) = 4.08$ ,  $p < .05$ ,  $\eta_p^2 = .127$ ). To clarify the MANOVA results, separate ANOVAs with the factors cueing (yes, no) and self-explaining (yes, no) were conducted for the performances on the retention, the inference and the transfer tests.

The ANOVA on the retention scores revealed no main effects of self-explaining ( $F(1,86) = 2.42$ ,  $MSE = 18.77$ ,  $p > .05$ ,  $\eta_p^2 = .027$ ) or cueing ( $F(1,86) = 0.30$ ,  $MSE = 18.77$ ,  $p > .05$ ,  $\eta_p^2 = .003$ ). However, the interaction between self-

Table 1. Means (and standard deviations) on the dependent measures as a function of condition

	Self-explaining				No self-explaining			
	Cued		Uncued		Cued		Uncued	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Learning outcomes								
Retention test (0–26)	11.38	4.41	8.71	4.01	10.64	4.26	12.30	4.61
Inference test (0–23)	10.33	2.54	6.52	3.67	4.52	2.63	5.09	3.15
Transfer test (0–21)	2.57	2.56	1.05	1.12	0.92	1.22	1.00	1.31
Mental effort (ME)								
ME Animation (1–9)	5.57	1.99	4.86	2.01	5.44	2.22	4.71	2.08
ME Inference test (1–9)	5.81	2.06	6.00	1.41	6.48	2.04	5.91	1.68
ME Retention test (1–9)	6.19	2.21	6.33	1.39	6.68	1.57	6.22	1.70
ME Transfer test (1–9)	5.91	1.94	6.48	2.02	6.16	1.93	6.00	2.63
Instructional efficiency (E)								
Retention	0.17	1.10	–0.32	0.82	–0.15	1.04	0.30	1.12
Inference	0.83	0.947	0.033	0.92	–0.53	1.02	–0.21	0.98
Transfer	0.57	1.17	–0.24	0.91	–0.19	0.86	–0.10	1.04

explaining and cueing was significant ( $F(1,86) = 5.59$ ,  $MSE = 18.77$ ,  $p < .05$ ,  $\eta_p^2 = .061$ ). *Post hoc* tests revealed no significant differences between any of the four conditions (all  $p$ 's  $> .05$ ).

The ANOVA on the inference scores showed a significant effect of self-explaining, with the explaining conditions making significantly more inferences than the no-explaining conditions ( $F(1,86) = 32.35$ ,  $MSE = 9.09$ ,  $p < .05$ ,  $\eta_p^2 = .273$ ). The effect of cueing was also significant ( $F(1,86) = 6.47$ ,  $MSE = 9.09$ ,  $p < .05$ ,  $\eta_p^2 = .070$ ), with a higher number of inferences for the cued than for the uncued conditions. The interaction between cueing and self-explaining was significant ( $F(1,86) = 11.79$ ,  $MSE = 9.09$ ,  $p < .05$ ,  $\eta_p^2 = .121$ ). *Post hoc* analyses showed that learners in the cued self-explaining condition generated significantly more inferences than learners in the uncued self-explaining condition ( $t(40) = 3.91$ ,  $p < .05$ ,  $d = 1.21$ ), the cued non-explaining condition ( $t(44) = 7.59$ ,  $p < .05$ ,  $d = 2.25$ ) and the uncued non-explaining condition ( $t(42) = 6.05$ ,  $p < .05$ ,  $d = 1.84$ ). The other comparisons revealed no significant differences. The interaction indicates that self-explaining yielded more correct inferences with than without spotlight-cues, whereas for the no-explaining conditions no significant difference was found between spotlight-cues and no spotlight-cues.

The ANOVA on the transfer scores showed a pattern analogous to the results on the inference test. A significant effect of self-explaining was found in the transfer test, indicating that learners in the explaining conditions obtained significantly higher scores than learners in the no-explaining conditions ( $F(1,86) = 6.04$ ,  $MSE = 2.67$ ,  $p < .05$ ,  $\eta_p^2 = .066$ ). In addition, there was a significant effect of cueing, with the cued conditions scoring significantly higher than the uncued conditions ( $F(1,86) = 4.36$ ,  $MSE = 2.67$ ,  $p < .05$ ,  $\eta_p^2 = .048$ ). The interaction between self-explaining and cueing was significant ( $F(1,86) = 5.38$ ,  $MSE = 2.67$ ,  $p < .05$ ,  $\eta_p^2 = .059$ ). The separate group means show a superiority of explaining in the cued condition, but not in the uncued condition. *Post hoc* analyses revealed that self-explaining with spotlight-cues resulted in significantly higher transfer scores than self-explaining without spotlight-cues ( $t(40) = 2.50$ ,  $p < .05$ ,  $d = .77$ ) or not self-explaining with cues ( $t(44) = 2.86$ ,  $p < .05$ ,  $d = .82$ ) and without cues ( $t(42) = 2.59$ ,  $p < .05$ ,  $d = .77$ ). The other comparisons revealed no significant differences.

A MANOVA on the mental effort invested in studying the animation and completing each knowledge test with cueing (yes, no) and self-explaining (yes, no) as the between-participants factors revealed no significant differences on either self-explaining (Wilks'  $\lambda = 0.98$ ,  $F(4,83) = 0.42$ ,  $p > .05$ ) or cueing (Wilks'  $\lambda = 0.95$ ,  $F(4,83) = 1.06$ ,  $p > .05$ ), nor a significant interaction between self-explaining and cueing was observed (Wilks'  $\lambda = 0.99$ ,  $F(4,83) = 0.31$ ,  $p > .05$ ). However, combining mental effort and performance measures may provide more valuable information on the quality of the constructed cognitive schema than performance measures alone (Van Gog & Paas, 2008). Therefore, test performance on the inference, retention and transfer tests and the mental effort invested during the performance of each of these tests were combined

using Paas and Van Merriënboer's (1993) computational approach to calculate instructional efficiency (E) on retention, inferences and transfer.

A MANOVA on the efficiency scores for retention, inference and transfer with cueing (yes, no) and self-explaining (yes, no) as the between-participants factors revealed a significant main effect for self-explaining (Wilks'  $\lambda = 0.74$ ,  $F(3, 84) = 10.02$ ,  $p < .05$ ,  $\eta_p^2 = .264$ ) and a significant interaction between self-explaining and cueing (Wilks'  $\lambda = 0.91$ ,  $F(3, 84) = 2.72$ ,  $p < .05$ ,  $\eta_p^2 = .088$ ). The main effect for cueing was not significant (Wilks'  $\lambda = 0.95$ ,  $F(3, 84) = 1.56$ ,  $p > .05$ ). Separate ANOVAs on the efficiency scores for retention, inference and transfer with cueing and self-explaining as the between-participant factors showed a significant effect of self-explaining on the efficiency scores of the inference test ( $F(1, 86) = 15.33$ ,  $p < .05$ ,  $\eta_p^2 = .151$ ), indicating that the explaining conditions were more efficient than the no-explaining conditions, that is, they obtained better performances on the inference test without investing more mental effort to obtain this performance. Moreover, there was a significant interaction between self-explaining and cueing for the efficiency scores on retention ( $F(1, 86) = 4.65$ ,  $p < .05$ ,  $\eta_p^2 = .051$ ), inference ( $F(1, 86) = 7.56$ ,  $p < .05$ ,  $\eta_p^2 = .081$ ) and transfer ( $F(1, 86) = 4.49$ ,  $p < .05$ ,  $\eta_p^2 = .050$ ). *Post hoc* analyses showed that the self-explaining with spotlight-cues condition was significantly more efficient on the inference test ( $F(3, 86) = 8.03$ ,  $p < .05$ ) and on the transfer test ( $F(3, 86) = 4.65$ ,  $p < .05$ ) than the other three conditions. The other comparisons revealed no significant differences. In short, the results on the efficiency scores largely parallel the results of the learning performance measures.

### Self-explanations during learning

To gather more information on how cues affected self-explaining during learning, the self-explanation protocols of participants in the two self-explaining conditions were analysed according to the coding scheme described above. Two independent raters who were unaware of experimental condition scored the verbal protocol data. One rater first transcribed the participants' utterances, separated them into statements and then categorized all statement into either a type of self-explanation statement, monitoring statement, incorrect self-explanation or paraphrase. A second rater recoded a randomly selected sample of 30% of the statements. To estimate interrater reliability, Cohen's Kappa ( $k$ ) was calculated on the pairs of scores in this subset. Results revealed a high interrater agreement ( $k = 0.88$ ), and therefore coding was considered sufficiently reliable. As participants' verbal protocols sometimes contained unintelligible or off-task utterances, the coded statements in the cued and uncued conditions covered respectively 97.8 and 97.3% of the words in the participants' verbal protocols.

The influence of cueing on the verbalizations was analysed by a MANOVA with cueing as the between-participants factor with number of words, number of self-explanation prompts and total number of statements as dependent variables. This analysis showed that cueing significantly influenced the characteristics of the learners' self-explanations (Wilks'

Table 2. Descriptive self-explanation data and mean number (and standard deviations) of statements per subsystem for the cued and uncued conditions

	Self-explaining			
	Cued		Uncued	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
General				
Word count	350.62	123.27	290.67	82.94
Self-explanation prompts	12.76	4.15	13.10	4.36
Total number of statements	25.24	6.02	20.86	4.51
Statements per subsystem				
Electrical system	3.52	1.37	3.05	2.01
Circulatory system	5.67	2.48	6.95	2.60
Systemic circulation	6.05	2.69	4.52	2.36
Valves system	5.76	2.68	2.62	2.16
Pulmonary circulation	4.24	1.73	3.14	2.08

$\lambda = 0.81$ ,  $F(3,38) = 2.95$ ,  $p < .05$ ,  $\eta_p^2 = .189$ ). More specifically, the results, depicted in Table 2, show that participants in the cued self-explanation condition did not use significantly more words in their protocols ( $F(1,40) = 3.42$ ,  $MSE = 11036.94$ ,  $p > .05$ ), and did not receive more self-explanation prompts from the experimenter ( $F(1,40) = 0.64$ ,  $MSE = 18.09$ ,  $p > .05$ ) than the uncued condition. However, learners in the cued condition made more statements about the animation than the uncued condition ( $F(1,40) = 9.10$ ,  $MSE = 28.30$ ,  $p < .05$ ,  $\eta_p^2 = .182$ ). So, on the descriptive level, only a significant difference was found between the cued and the uncued self-explanation conditions on the number of statements about the animation.

To gain more insight into the content of the statements, we looked at what subsystems of the animation were explained by participants (Table 2). A MANOVA with cueing as between-participants factor and total number of statements (i.e. all paraphrases, goal-directed explanations, elaborative explanations, incorrect explanations together) about each subsystem as dependent variables showed a cueing-effect (Wilks'  $\lambda = 0.55$ ,  $F(5,36) = 6.01$ ,  $p < .05$ ,  $\eta_p^2 = .953$ ). *Post hoc* analyses revealed that more statements were made about the valves system in the cued than in the uncued condition ( $F(1,40) = 17.52$ ,  $MSE = 5.92$ ,  $p < .05$ ,  $\eta_p^2 = .305$ ). The other comparisons showed no significant differences. So, cues only led to a higher number of statements for one out of five subsystems.

Furthermore, in order to examine to what extent cueing influenced the type of statements learners made about the animation we conducted a MANOVA with cueing as between-participants factor and paraphrases, goal-directed self-explanations, elaborative self-explanations, monitoring statements, incorrect self-explanations as dependent variables. It should be noted that monitoring statements only contained negative utterances as no positive monitoring statements were made. Results, shown in Table 3, revealed that cueing influenced what types of statements were generated (Wilks'  $\lambda = 0.44$ ,  $F(5,34) = 9.12$ ,  $p < .05$ ,  $\eta_p^2 = .559$ ). There were no significant differences in the number of paraphrases ( $F(1,40) = 0.11$ ,  $MSE = 17.37$ ,  $p > .05$ ). However, self-explaining with spotlight-cues elicited significantly more elaborative ( $F(1,40) = 24.95$ ,  $MSE = 6.57$ ,  $p < .05$ ,  $\eta_p^2 = .384$ ) and goal-directed self-explanations ( $F(1,40) = 31.39$ ,  $MSE = 2.13$ ,  $p < .05$ ,

$\eta_p^2 = .440$ ) than self-explaining without spotlight-cues. In addition, participants who self-explained with cues generated fewer (negative) monitoring statements ( $F(1,40) = 9.43$ ,  $MSE = 1.46$ ,  $p < .05$ ,  $\eta_p^2 = .191$ ) and incorrect self-explanations ( $F(1,40) = 4.72$ ,  $MSE = 1.14$ ,  $p < .05$ ,  $\eta_p^2 = .105$ ) than learners who self-explained without cues. So, cueing influenced what types of statements learners produced during self-explaining.

As a further refinement, we looked at the type of statement (i.e. paraphrase, goal-directed explanation, elaborative explanation, incorrect explanation) focused on each subsystem. We therefore conducted separate ANOVA's on the statements about each subsystem with cueing as between-participants factor and type of statement as within-participants factor. This allowed us to look at the interaction between cueing and type of statement. Results, depicted in Table 3, showed a significant interaction between cueing and type of statement for the number of paraphrases per subsystem ( $F(4,160) = 5.94$ ,  $MSE = 3.85$ ,  $p < .05$ ,  $\eta_p^2 = .129$ ), but there was no significant effect of cueing on the number of paraphrases per subsystem ( $F(1,40) = 0.53$ ,  $MSE = 3.94$ ,  $p > .05$ ). So, it seems that cueing led to a different number of paraphrases on each subsystem. Furthermore, there was no significant interaction for both goal-directed ( $F(4,160) = 0.80$ ,  $MSE = .44$ ,  $p > .05$ ) and elaborated explanations ( $F(4,160) = 1.34$ ,  $MSE = .72$ ,  $p > .05$ ) focused on each subsystem. Nevertheless, the effect of cueing was significant for the number of goal-directed ( $F(1,40) = 31.39$ ,  $MSE = 0.44$ ,  $p < .05$ ,  $\eta_p^2 = .440$ ) and elaborated explanations ( $F(1,40) = 24.95$ ,  $MSE = 1.32$ ,  $p < .05$ ,  $\eta_p^2 = .384$ ) per subsystem. This suggests that over all subsystems the number of relevant types of self-explanations is higher in the cued compared to the uncued condition. No significant interaction ( $F(4,160) = 1.35$ ,  $MSE = .20$ ,  $p > .05$ ) and no main effect of cueing ( $F(1,40) = 2.87$ ,  $MSE = 0.24$ ,  $p > .05$ ) was found on the number of incorrect self-explanations, suggesting that irrespective of cueing equal numbers of incorrect explanations were made per subsystem.

Furthermore, as sequentially presented cues might help organize learners' self-explanations, we analysed the extent to which cueing ordered the generation of self-explanations over the subsystems. If learners link their statements to

Table 3. Mean number (and standard deviations) of type of statement over all subsystems and specified for each subsystem for the cued and uncued conditions

Type of statement	Self-explaining			
	Cued		Uncued	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Paraphrase	13.86	4.36	14.29	3.69
Goal-driven self-explanation	3.86	1.80	1.33	1.02
Elaborative self-explanation	6.95	3.06	3.00	1.95
Monitoring	0.81	.087	1.95	1.47
Incorrect self-explanation	0.71	0.64	1.42	1.36
Type of statement per subsystem				
Paraphrase—electrical system	1.19	0.93	1.62	1.53
Paraphrase—circulatory system	2.43	2.01	4.81	2.48
Paraphrase—systemic circulation	3.76	2.21	3.48	2.44
Paraphrase—valves system	3.71	2.28	2.00	1.95
Paraphrase—pulmonary circulation	2.43	1.54	2.62	1.75
Goal-directed—electrical system	1.10	1.04	0.52	0.75
Goal-directed—circulatory system	0.81	0.75	0.43	0.68
Goal-directed—systemic circulation	0.48	0.60	0.05	0.22
Goal-directed—valves system	1.05	0.67	0.24	0.54
Goal-directed—pulmonary circulation	0.48	0.68	0.10	0.30
Elaborative—electrical system	0.57	0.51	0.19	0.40
Elaborative—circulatory system	2.43	1.25	1.33	1.20
Elaborative—systemic circulation	1.81	1.12	0.91	0.94
Elaborative—valves system	0.81	0.87	0.24	0.44
Elaborative—pulmonary circulation	1.33	1.24	0.33	0.58
Incorrect—electrical system	0.67	0.66	0.71	0.90
Incorrect—circulatory system	0.00	0.00	0.38	0.59
Incorrect—systemic circulation	0.00	0.00	0.10	0.30
Incorrect—valves system	0.19	0.40	0.14	0.36
Incorrect—pulmonary circulation	0.00	0.00	0.10	0.30
Relative number of switches between subsystems	0.29	0.09	0.36	0.11

the cues, we would expect fewer subsystem switches in their statements than when there are no cues in the animation. We therefore looked at the relative number of times that learners' statements contained a switch from one subsystem to the other as an indicator of ordered processing. Results, shown in Table 3, were consistent with our hypothesized idea, showing that the relative number of switches between subsystems was significantly lower in the cued condition than in the uncued condition ( $t(40) = 2.09$ ,  $p < .05$ ,  $d = .10$ ).

To examine how the self-explanation data were related to the performance on the learning outcome measures, Pearson Correlation Coefficients were calculated for each type of self-explanation participants generated and their retention, inference and transfer scores (Table 4). Both goal-directed

self-explanations and elaborative self-explanations were positively related to the scores on the inference test, whereas goal-directed self-explanations also related positively to retention test scores and transfer test scores. A higher number of self-explanation errors was associated with lower scores on the inference test. Further, fewer paraphrases were associated with higher transfer test scores.

## DISCUSSION

This study investigated the effects of generating self-explanations on learning from a complex animation with or without spotlight-cues. It was hypothesized that learning performance when studying a cued animation would be

Table 4. Interrelations between the types of self-explanation statements and learning outcome measures

	1	2	3	4	5	6	7	8
(1) Retention test	—	0.66**	0.49**	-0.07	0.40**	0.28	-0.14	-0.23
(2) Inference test		—	0.57**	-0.06	0.60**	0.34*	-0.29	-0.31*
(3) Transfer test			—	-0.35*	0.51**	0.19	-0.22	-0.17
(4) Paraphrases				—	-0.16	-0.11	-0.03	0.23
(5) Goal-directed self-explanation					—	0.55**	-0.40**	-0.33*
(6) Elaborative self-explanation						—	-0.35*	-0.36*
(7) Monitoring							—	0.10
(8) Incorrect self-explanation								—

\* $p < 0.05$ ; \*\* $p < 0.01$ .

enhanced by generating self-explanations, whereas self-explaining with an uncued animation would not enhance learning performance. As expected, the results showed that learners who generated self-explanations with a cued animation yielded higher performance on inference and transfer (but not retention) tasks than learners who generated self-explanations with an uncued animation and learners who did not self-explain. Thus, test performance of learners in the cued self-explanation condition indicates that they had developed a more thorough conceptual understanding of the causal relations in the cardiovascular system. Hence, they were better able to apply the learned information to novel tasks. Moreover, the cued self-explanation condition seems most efficient for learning, that is, learners had a higher performance without investing more mental effort to obtain this performance.

These findings provide new evidence regarding the effectiveness of cueing in animations by showing that attention-directing cues can foster learning from an animation if learners are simultaneously cognitively stimulated. If not, cues do not support learning, which is consistent with the increasing evidence that merely emphasizing individual locations in animations with cues does not foster learners' understanding of animations (for an overview, see De Koning, Tabbers, Rikers, & Paas, 2009).

Additionally, our findings extend the substantial literature showing that self-explaining is an effective strategy to improve learners' understanding of texts (Chi *et al.*, 1994), diagrams (Ainsworth & Loizou, 2003), or the simultaneous presentation of text and static graphics (Butcher, 2006). This study demonstrates that prompting learners to self-explain improves their understanding of an instructional animation without textual explanation. In contrast to the majority of self-explanation studies where learners engage in elaborate training and practice of self-explaining (McNamara, 2004), our study shows that even with minimal training and limited time for self-explaining, significant learning gains can be obtained. Whether this dissimilarity is due to differences in processing demands of textual and graphical representations, is an issue for future research.

Further, the present findings extend the results of prior animation studies showing that *a priori* encouraging learners to invest more mental effort in relevant learning activities by prequestions or having learners predict the next step in a process leads to better learning performance (Hegarty *et al.*, 2003; Mayer, Dow, & Mayer, 2003). This study demonstrates that *directly* prompting learners to explain their thoughts while studying the animation improves their understanding, at least when the animation is cued.

The availability of attention-directing cues appears to influence the effectiveness of self-explanations and hence construct a deeper understanding. As predicted, although learners in both self-explanation conditions generated adequate self-explanations, the animation with cues increased the frequency with which learners generated goal-driven and elaborative self-explanations (Ainsworth & Loizou, 2003). Further, the results suggest that cues prevented learners from generating incorrect inferences. The finding that certain types of active processing and hence the type of self-explanation statements largely influence

learning performance is consistent with prior research on self-explaining in text and/or diagrams (Ainsworth & Burcham, 2007; Butcher, 2006).

The question remains, however, why cueing led to more relevant and correct self-explanations? As both self-explanation conditions did not differ in the number of statements about the subsystems (except for the valves), it does not seem likely that the uncued group missed much information or did focus on other parts of the animation than the cued group. This is in line with prior research that cues do not alter learners' total distribution of attention over the subsystems (De Koning *et al.*, in press). Rather, as learners in the cued condition generated a higher number of meaningful self-explanations for all subsystems, it seems that cues enabled learners to process information more deeply. Moreover, the fact that in the cued condition the statements about the subsystems were more grouped together than in the uncued condition, as indicated by fewer subsystem switches, suggests that cueing helped learners to better organize their self-explanations. Following Butcher (2006) that highlighting essential parts in visualizations leads to more relevant types of self-explanations, our findings seem to suggest that showing different subsystems of the heart consecutively in the cued animation may have provided learners a structured sequence of meaningful functional units, which helped them focus on and explain each part of the animation in more depth and for a certain amount of time. In contrast, in the uncued condition the explanations seemed less focused and grounded on the subsystems, thereby hindering deeper processing of each part of the animation. Further research identifying the exact role of attention guidance, information structuring or factors like motivation could further develop this explanation.

The amount of cognitive load may also have influenced learning from cued and uncued animations. Several studies have provided explanations for cueing-effects by suggesting that cues reduce ineffective working memory load and increase cognitive resources for learning-related activities or germane load (Paas, Van Gerven, & Wouters, 2007). In our study then learners should use more mental resources for providing self-explanations (i.e. germane load) in the cued than the uncued condition. However, as our mental effort measure did not differentiate between different types of cognitive load, we cannot draw definite conclusions in this regard. Based on the observation that the generation of important types of self-explanations was best in the cued condition and that learning in this condition was most efficient (i.e. higher test performance without differences in mental effort), we can only assume that if self-explaining with a cued animation imposes extra cognitive load it is likely to be related to learning (i.e. germane load). Future studies are needed to test such capacity-related explanations.

In the following we discuss a number of objections to our findings as well as avenues for future research. It might be argued that the superior performance with self-explaining is a bit overstated because of the dissimilarity between the individual testing in the self-explanation conditions and the group testing that sometimes occurred in the no self-

explanation conditions. However, although there was a significant main effect for self-explanation, the effect was primarily explained by the significant interaction between cueing and self-explaining, which indicated better test performances for the cued self-explaining group, but not for the uncued self-explaining group. Furthermore, although non-explainers did perform worse than self-explainers on the inference and transfer tests, their performance on the retention test was not, nor did they report lower mental effort scores. So, although we cannot exclude the possibility of confounding of test performance in the self-explanation condition, we believe this is highly unlikely.

Furthermore, it is yet unclear what role the prompts have played and to what extent they were necessary in obtaining the current results. Cued and uncued self-explainers did not receive different amounts of prompts and prompts alone seemed insufficient to improve learning. However, as Chi et al. (1994) have pointed out, self-explanation prompts are very helpful but may not be necessary for improving learning. Future research should clarify this issue for our results.

Another interesting observation is that in all conditions transfer scores were low, which is likely to be due to the high difficulty of the questions, requiring learners to go through several reasoning steps about the functioning of the heart in order to arrive at a correct solution. It is possible that transfer items requiring fewer reasoning steps would have resulted in better transfer performances. Further research in which all participants are tested individually, prompted and non-prompted groups are compared, and appropriate transfer tests are used, is needed to warrant any definite claims on these issues and the implications concerning the generalizability of the findings.

In contrast to the majority of animation studies conducted in a lab setting, the current study was an attempt to investigate perceptual and cognitive manipulations in learning from an animation in a classroom setting (also see e.g. Marbach-Ad, Rotbain, & Stavy, 2008). The use of a non-narrated and non-interactive animation and the absence of supplementary materials such as introductory texts made it possible to exert, to a large extent, control over the manipulated factors in our study. However, we are aware that this approach does not exactly correspond to a regular classroom situation and largely constrains the generalizability of the findings for real teaching situations. Therefore, a replication of the experiment in a real learning environment in which the animation is integrated in the curriculum and where is less control over its features would be needed to further strengthen our conclusions. Nevertheless, we believe that controlled classroom studies provide a necessary and important step towards investigating learning from animations embedded in a real learning situation. These kinds of studies (including the present study) may identify specific conditions under which learning from animations may be enhanced that subsequently can be tested in existing educational programmes.

Moreover, further research is required to examine whether the present findings can be generalized more broadly to longer educational tasks, to different types of animations (e.g. non-cyclical), using different types of cueing (e.g.

arrows), and applying other active processing strategies (e.g. asking questions). Specifically, the type of animation used in this study might have been a major determinant in finding the observed effects. The animation's cyclic nature may have facilitated self-explaining because learners could iteratively view—individual parts of—the animation and hence could monitor their current understanding and update their mental representation accordingly. It is unclear whether viewing a non-cyclical animation, in which information cannot be re-inspected and thus might be easily missed or partly processed, will have similar effects on learning. In addition, it would be interesting to investigate whether providing interactivity (e.g. rewind, stop) may lead to different learning performances when self-explaining with cued or uncued animations. Especially, investigating the combination of cueing and learner control would be relevant because recent research has demonstrated that providing learners with control over an animation does not necessarily improve learning, whereas seeing the most task-relevant information does, regardless of whether task-relevant information was obtained with or without learner control (Keehner, Hegarty, Cohen, Khooshabeh, & Montello, 2008). Furthermore, presenting the animation without textual explanations may have provided a productive situation for enhancing conceptual understanding with self-explaining. Generating verbal self-explanations might improve understanding with a visual-only animation as it encourages learners to integrate visual (i.e. animation) and verbal (i.e. self-explanations) representations. In contrast, when instructions and self-explanations use the same form of representation, such as self-explaining with a narrated animation, self-explanations and narrations are redundant, which is likely to increase extraneous load and hence self-explaining will be less effective at promoting deeper understanding (Cox, 1999). Future studies should therefore focus on how various attention-guiding techniques and learning strategies influence learning from different types of animations.

In sum, this study suggests that presenting learners with a complex instructional animation without cues and stimulating them to engage in cognitive activities is insufficient to obtain improved learning performances. Effective learning requires that learners are supported by attention-directing cues in selecting information that will subsequently serve as the basis for essential cognitive processing. Therefore, a practical implication resulting from this study is that we should design animations that support the processing of critical information to effectively verbalize the operation of a dynamic system and to foster learning. Self-explaining can be an effective tool to enhance learning from a complex animation without text provided that the animation is designed to highlight and structure essential components and their (causal) relations that are necessary to understand the depicted system.

#### ACKNOWLEDGEMENTS

The thank the students and staff of the Thorbecke Lyceum in Rotterdam, The Netherlands for their participation in this study.

## REFERENCES

- Ainsworth, S., & Burcham, S. (2007). The impact of text coherence on learning by self-explanation. *Learning and Instruction, 17*, 286–303.
- Ainsworth, S., & Loizou, A. T. (2003). The effects of self-explaining when learning with text or diagrams. *Cognitive Science, 27*, 669–681.
- Ainsworth, S., & VanLabeke, N. (2004). Multiple forms of dynamic representation. *Learning and Instruction, 14*, 214–255.
- Aleven, V., & Koedinger, K. R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based cognitive tutor. *Cognitive Science, 26*, 147–179.
- Ayres, P., & Paas, F. (2007). Making instructional animations more effective: A cognitive load approach. *Applied Cognitive Psychology, 21*, 695–700.
- Bétrancourt, M. (2005). The animation and interactivity principles in multimedia learning. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 287–296). New York: Cambridge University Press.
- Butcher, K. R. (2006). Learning from text with diagrams: Promoting mental model development and inference generation. *Journal of Educational Psychology, 98*, 182–197.
- Chi, M. T. H., de Leeuw, N., Chiu, M. H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*, 439–477.
- Chi, M. T. H., Siler, S. A., Jeong, H., Yamauchi, T., & Hausmann, R. G. (2001). Learning from human tutoring. *Cognitive Science, 25*, 471–533.
- Cox, R. (1999). Representation construction, externalised cognition and individual differences. *Learning and Instruction, 9*, 343–363.
- De Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2007). Attention cueing as a means to enhance learning from an animation. *Applied Cognitive Psychology, 21*, 731–746.
- De Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2009). Towards a framework for attention cueing in instructional animations: Guidelines for research and design. *Educational Psychology Review, 21*, 113–140.
- De Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (in press). Attention guidance in learning from a complex animation: Seeing is understanding? *Learning and Instruction*. DOI: 10.1016/j.learninstruc.2009.02.010
- Hegarty, M. (2005). Multimedia learning about physical systems. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 447–465). New York: Cambridge University Press.
- 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.
- Hillstrom, A. P., & Chai, Y. C. (2006). Factors that guide or disrupt attentive visual processing. *Computers in Human Behavior, 22*, 648–656.
- Höfler, T., & Leutner, D. (2007). Instructional animation versus static pictures: A meta-analysis. *Learning and Instruction, 17*, 722–738.
- Keehner, M., Hegarty, M., Cohen, C., Khooshabeh, P., & Montello, D. R. (2008). Spatial reasoning with external visualizations: What matters is what you see, not whether you interact. *Cognitive Science, 32*, 1099–1132.
- Kriz, S., & Hegarty, M. (2007). Top-down and bottom-up influences on learning from animations. *International Journal of Human-Computer Studies, 65*, 911–930.
- Lowe, R. K. (1999). Extracting information from an animation during complex visual learning. *European Journal of Psychology of Education, 14*, 225–244.
- Lowe, R. K. (2003). Animation and learning: selective processing of information in dynamic graphics. *Learning and Instruction, 13*, 157–176.
- Macromedia. (2004). *Flash MX 7.0 [Computer software]*. San Francisco: Author.
- Marbach-Ad, G., Rotbain, Y., & Stavy, R. (2008). Using computer animation and illustration activities to improve high school students' achievement in molecular genetics. *Journal of Research in Science Teaching, 45*, 273–292.
- Mayer, R. E. (2001). *Multimedia learning*. New York: Cambridge University Press.
- Mayer, R. E., Dow, G. T., & Mayer, S. (2003). Multimedia learning in an interactive self-explaining environment: What works in the design of agent-based microworlds? *Journal of Educational Psychology, 95*, 806–812.
- Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology, 90*, 312–320.
- McNamara, D. S. (2004). SERT: Self-explanation reading training. *Discourse Processes, 38*, 1–30.
- 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.
- Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38*, 1–4.
- Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*, 63–71.
- Paas, F., Van Gerven, P. W. M., & Wouters, P. (2007). Instructional efficiency of animation: Effects of interactivity through mental reconstruction of static key frames. *Applied Cognitive Psychology, 21*, 783–793.
- Paas, F., & Van Merriënboer, J. J. G. (1993). The efficiency of instructional conditions: An approach to combine mental-effort and performance measures. *Human Factors, 35*, 737–743.
- Peterson, L., & Peterson, M. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology, 58*, 193–198.
- Renkl, A. (2005). The worked-out-example principle in multimedia learning. In R. Mayer (Ed.), *Cambridge handbook of multimedia learning*. Cambridge, UK: Cambridge University Press.
- Renkl, A., & Atkinson, R. K. (2002). Learning from examples: Fostering self-explanations in computer-based learning environments. *Interactive Learning Environments, 10*, 105–119.
- Renkl, A., & Atkinson, R. K. (2003). Structuring the transition from example study to problem solving in cognitive skills acquisition: A cognitive load perspective. *Educational Psychologist, 38*, 15–22.
- Renkl, A., Stark, R., Gruber, G., & Mandl, H. (1998). Learning from worked-out examples: The effects of example variability and elicited self-explanations. *Contemporary Educational Psychology, 23*, 90–108.
- Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science, 26*, 521–562.
- Tversky, B., Morrison, J. B., & Bétrancourt, M. (2002). Animation: Can it facilitate? *International Journal of Human-Computer Studies, 57*, 247–262.
- Van Gog, T., & Paas, F. (2008). Instructional efficiency: Revisiting the original construct in educational research. *Educational Psychologist, 43*, 16–26.