Original article

doi: 10.1111/jcal.12136

Journal of Computer Assisted Learning

The effects of reviews in video tutorials

H. van der Meij* & J. van der Meij†

*Faculty of Behavioural, Management and Social Sciences, Department of Instructional Technology, University of Twente, Netherlands
†Faculty of Behavioural, Management and Social sciences, ELAN. Institute for Teacher Education and Science Communication, University of Twente, Netherlands

Abstract

This study investigates how well a video tutorial for software training that is based on Demonstration-Based Teaching supports user motivation and performance. In addition, it is studied whether reviews significantly contribute to these measures. The Control condition employs a tutorial with instructional features added to a dynamic task demonstration. The Review condition additionally includes video reviews. Participants were 55 seventh graders who viewed task demonstrations (and reviews) followed by practice. Both tutorials increased motivation (i.e., task relevance and self-efficacy) and performance. In addition, the Review condition had significantly better results for training time, self-efficacy and scores on an immediate post-test. Reviews have rarely been studied in dynamic visualizations. The present study suggests that there may be important advantages to be gained from concluding a demonstration video with a summary of the main points.

Keywords Demonstration-Based Training, reviews, software tutorial, video.

Introduction

Until recently, most of the instructional support for the beginning and moderate software user came from paper tutorials. Under the influence of YouTube's rapid growth in popularity, and supported by easy to use programmes for video production, editing and sharing, more and more software companies and third party vendors have begun to switch to video as the primary medium for their tutorials (van der Meij, Karreman, & Steehouder, 2009). This raises the question whether video tutorials for software training can be at least as effective as the paper tutorials that they are replacing. This issue brings us into the current discussion on the advantages of dynamic versus static visualizations, a debate that revolves around the critical boundary conditions (e.g., Lowe, Schnotz, & Rasch, 2011; Brucker, Scheiter, & Gerjets, 2014). The key question is when one or the other form can be expected to be more effective.

The most important criterion is that the depiction should be aligned with the type of mental representation required of the user. Dynamic representations such as video can be expected to benefit the user only when there is a fit between the content and structure of what the user sees and what must be remembered. In software training the aim is the acquisition of procedural knowledge. The user must get to know the sequence of steps that lead to task completion in a particular software programme. This requires the user to learn to perform a series of actions that lead to changes on the screen that must be observed. According to the congruence principle, dynamic visualizations (e.g., video) should be particularly beneficial for learning such a task (Tversky, Bauer-Morrison, & Bétrancourt, 2002).

A suitable means of instructing people about a procedure comes from demonstrating task performance (Smith & Ragan, 2005). In order to learn from such a demonstration, the user must carefully observe the modelled procedure. Bandura's (1986) social–cognitive learning theory

Accepted: 10 January 2016
Correspondence: Hans van der Meij, University of Twente, Faculty of Behavioural, Management and Social sciences, Department of Instructional Technology, Drienerlolaan 5, 7522 AE Enschede, the Netherlands. Email: h.vandermeij@utwente.nl

provides fundamental insights in the processes involved in such observational learning. According to this theory, learning from task demonstrations involves the interrelated processes of attention, retention, production and motivation. The construction of a video should cater for these processes.

The next section describes the four processes and design measures to support these. Special attention is given to retention for which the study investigates a unique design measure, namely the inclusion of a review video. Just as with end summaries in paper texts, it is expected that a concise summary of task achievement after a demonstration contributes to the user's memory of a procedure. The remainder of the paper reports on an experiment in which the effectiveness of a video tutorial with demonstrations is compared with a video tutorial with demonstrations plus reviews.

Demonstration-Based Training

There is an extensive literature on the design and effectiveness of demonstration videos for motor skills development (e.g., Schwan & Riempp, 2004; Ayres, Marcus, Chan, & Qian, 2009; Akinlofa, O'Brian Holt, & Elyan, 2013) and problem solving (e.g., Spanjers, van Gog, Wouters, & van Merrienboer, 2012; Hoogerheide, Loyens, & van Gog, 2014). This literature is only partly relevant for the design of a video tutorial for software training, however, because of a difference in goals. Motor skills training primarily revolves around learning physical actions (e.g., hand movements). In contrast, the emphasis in software training lies on getting to know the software interface and learning the action-reaction patterns in its handling. The user must learn to apply a procedure on the interface, rather than learn how to act on an input device. In addition, task procedures are strictly defined. All steps must be included and each step is unambiguous. Therefore, procedures are sometimes qualified as algorithms (Smith & Ragan, 2005). This distinguishes procedures from problem solving which is more heuristic in nature.

For the construction of the video tutorial in this study we therefore also looked at two other sources. One source was recent research on the design and effectiveness of software training with video (e.g., Lloyd & Robertson, 2012; van der Meij & van der Meij, 2013; van der Meij & van der Meij, 2014; van der Meij & van der Meij, 2015). The other source was research on observational learning. Bandura (1986) draws attention to the basic processes that should be supported in model-based learning: attention, retention, production and motivation.

We discuss these processes next and complement their descriptions with the design guidelines that were followed in the construction of the video tutorial in this study. This design approach, in which task demonstrations are coupled with instructional measures for promoting learning, is called Demonstration-Based-Training (Rosen et al., 2010; Grossman, Salas, Pavlas, & Rosen, 2013).

Attention is an active process in which the demonstrated information is filtered or selected. Users must attend primarily to the salient information; they must concentrate on what is pertinent for task accomplishment and ignore other information. This process is made difficult by a combination of a complex user-interface and the transient nature of videos. The interface challenges the user to discover where and what to look for on the screen. The medium challenges the user to do so with continuous and rapid screen changes. This is a daunting task that calls for design measures that can support the user's attentional processes.

Two important design features that address the user's distribution of attentional resources are signalling and pacing.

A well-known design measure for directing attention in reading from text is signalling (Lemarié, Lorch, Eyrolle, & Virbel, 2008). In videos, two prevalent ways of supporting the user in allocating attention to pertinent screen information are highlighting and zooming. Both techniques draw the user's attention to the relevant place or object on the screen (van der Meij & van der Meij, 2013). Two recent empirical studies show that signalling techniques can positively affect learning from dynamic visualizations (Amadieu, Mariné, & Laimay, 2011; Jin, 2013).

Pacing is a slightly elusive design feature. The advice is that the pace should be moderate; it should not be too slow for risk of boredom, nor should it be too fast for risk of cognitive overload (Koumi, 2013). In short, the native pacing of the video should be adapted to what the audience can handle. Because it is hard to establish pace on the basis of design guidelines alone, it is best to pilot test the video for its pacing. A corollary design feature of pacing is the inclusion of a toolbar that the user can employ to play, rewind, pause or stop the video. Such a toolbar enables the user to adapt the pace of the video depending on what is needed to process the information. In their study on videos for learning to tie nautical knots, Schwan and Riempp (2004) reported important advantages of user-pacing. Among others, they found that users made a heavier use of toolbar functions to adapt the pacing of the videos on more difficult knots. More generally, user-pacing appears to be an important design feature that affects learning from dynamic visualizations (e.g., Stiller, Freitag, Zinnbauer, & Freitag, 2009; Witteman & Segers, 2010; Merkt, Weigand, Heier, & Schwan, 2011).

Retention refers to the comprehension and storing of information for future behaviour. The demonstration should be designed in such a way that the user can understand how a task is performed, and it should support the user in remembering the procedure so that it can serve as a guide for future action. Three main measures for supporting retention are optimized segment length, simple-to-complex task sequencing and the inclusion of pauses.

Presenting tasks in manageable units or segments facilitates the user's understanding of a procedure. Complex or long tasks should therefore be split into smaller units or segments. Such splits are preferably based on a meaningful subtask division. An important boundary condition to keep in mind in this respect is video length. Research suggests that a maximum length of 3 min is acceptable, but that a duration of 1 min is best to keep all users aboard (Plaisant & Shneiderman, 2005; Wistia, 2012; Guo, Kim, & Rubin, 2014).

Another way of supporting understanding comes from organizing the tasks in a simple-to-complex sequence. Placing easier tasks before more difficult ones has the advantage that the user can keep up with increasing levels of task complexity. On each moment in training the user then faces a task that should be manageable (van Merriënboer, Kirschner, & Kester, 2003).

The user can be supported in remembering a procedure by the inclusion of brief, 2- to 5-s pauses at key moments in a demonstration. A recent empirical study by Spanjers et al. (2012) shows that such pauses can support retention of dynamic representations in two ways. One, pauses can demarcate key units or segments for the user and thereby contribute to understanding. The breaks signal the important units or building blocks of which a procedure consists. Two, pauses interrupt the continuous stream of information in a dynamic presentation. The user can benefit from such a break by engaging in maintaining activities. The pause gives the user time to engage in mental rehearsal (Rosen et al., 2010).

Production refers to the learner's actions taken to accomplish the modelled task performance. The main instructional feature advocated for supporting this process is the inclusion of complementary practice (Grossman et al., 2013; van der Meij & van der Meij, 2013).

Practice can serve as a check of understanding and recall. During practice the user may come to realize that a step in the procedure is forgotten or an error is made leading to a need to restudy the video. This suggests that it is beneficial for the user to have easy access to the videos during practice. Practice can also consolidate a procedure. It can reinforce what the user remembers. Empirical studies on multimedia show that users usually benefits from practice, but that it may depend on their prior knowledge whether practice best occurs before or after a demonstration (e.g., Reisslein, Atkinson, Seeling, & Reisslein, 2006; Wouters, Paas, & van Merriënboer, 2010).

Motivation refers to the intensity, valence and persistence of one's learning-directed behaviour (Pintrich & Schunk, 2002). It is the driving force behind the processes of attention, retention and production. Earlier we mentioned the simple-to-complex sequencing of tasks as a facilitator of understanding. In addition, this design measure is likely to contribute to user motivation. Other features that can positively affect motivation are a task-oriented organization and the presence of a human narrator using a conversational style.

For software instructions, a distinction is often made between a function and a task orientation. The first refers to a presentation mode that concentrates on affordances. The user receives explanations of software functions, features and interface elements. Reference guides are sometimes organized in this fashion. In a task-oriented presentation, usage of the software by the audience is given a central role. The focus lies on selecting or creating tasks that the user instantly recognizes as genuine and meaningful (van der Meij & Carroll, 1998).

Another feature that can contribute to user motivation is the presence of a human voice that addresses the user in a conversational rather than formal style. Various empirical studies have found proof of this personalization effect (Kartal, 2010; Reichelt, Kämmerer, Niegemann, & Zander, 2014). A recent meta-study further substantiated this effect, noting that training time was an important moderator. When instructions take longer than 35 min the personalization effect disappeared (Ginns, Martin, & Marsh, 2013).

Summaries with text and video

Reviews, summaries of steps for task completion, can bring the principal solution steps that lead to task accomplishment back into the user's active memory. Thus, they would seem optimally suited to contribute to retention of the information needed to complete the task. Surprisingly, very little, if any, documentation exists for the design and effectiveness of video reviews. A literature search for empirical studies on the effectiveness of reviews in videos revealed no hits. The meta-analysis of expository animations by Ploetzner and Lowe (2012) also did not report a single case involving reviews among the 44 empirical studies that were analysed. When our search was extended to include summaries with texts the pursuit was only slightly more fruitful. Only a few older studies were discovered. Their findings are reported in the succeeding texts. Our discussion here concentrates on the studies that investigated summaries after a text, which we will call 'end summaries'.

Hartley, Goldie and Steen (1976) conducted an experiment in which they examined the influence of summary placement on text retention as indicated by recall. There were three conditions: (a) beginning summary, (b) end summary and (c) no summary. After reading the text, participants were asked a number of questions about the text. Recall was best for the ending summary. No differences were found between the beginning summary and the no summary conditions.

McLaughlin Cook (1981) considered the argument that the finding by Hartley et al. (1976) might be related to the attention paid to the summaries. That is, the absence of a positive effect for the beginning summary might be caused by readers skipping over it. To investigate this possibility, he conducted an experiment with four conditions: (a) beginning summary on the same page as the text ('beginning summary – same page'), (b) beginning summary on a separate page from the text ('beginning summary – separate page'), (c) end summary and (d) no summary. Text recall was measured with a set of questions. No difference in recall was found between the beginning summary – separate page and 'end summary' conditions, which both yielded significantly higher recall than the other conditions. The conclusion was that summaries could increase recall if their design is such that it stimulates readers to actively process them.

Hartley and Trueman (1982) reviewed the outcomes from the research conducted to that date. Besides the two previously mentioned studies, they reported having found one study from 1955 (i.e., Christensen & Stordahl) with no reliable effects, and one study from 1973 (i.e., Vezin, Berge, & Mavrelis) that reported a significant benefit from the inclusion of an end summary. Next, Hartley and Trueman gave an account of five consecutive empirical studies on the effect of the placement of summaries on text retention and recall. Four of the five experiments used texts that included (a) a beginning summary, (b) end summary or (c) no summary. Students were instructed to read a text (and summary) in order to make a judgment as to its readability. After reading the text (and summary) once, the students answered recall questions. The overall finding was that summaries consistently improved recall for cued information. There was no effect on recall of information not mentioned in the summary. Also, no difference was found for summary position. Beginning and end summaries were found to be equally supportive.

All in all, the few empirical studies on end summaries with texts show that these enhance retention and recall. The relative dearth of studies on the effectiveness of end summaries supports the contention of Hartley and Davies (1976) that their value 'seems to be so obvious that few people have felt any real need to subject the concept to empirical investigation' (p. 251).

Experimental design and research questions

This study compares a video tutorial on Word's formatting options that includes task demonstrations with and without reviews (Review and Control condition, respectively). Testing in software training usually revolves around the three main facets mentioned in ISO-9421, namely, engagement, effectiveness and efficiency. This study investigates these facets. More specifically, the following research questions are addressed:

Research question 1: *Does condition influence training time*?

Training time is checked to determine whether it is affected by the added presence of reviews. Because the review videos are very short, no difference between conditions is expected.

Research question 2: *How well do the video tutorials support motivation, and is there an effect of condition?*

The video tutorials are designed to make the training tasks meaningful and doable in participants' eyes. To assess their motivational impact, we examine task relevance and self-efficacy. Task relevance refers to present and future value of completing a task. It indicates the importance of a task to someone's goals or concerns (Pintrich & Schunk, 2002). Self-efficacy can be defined as a person's expectancy for success in novel tasks (Bandura, 1997). The two motivational constructs are measured before and after training. No effect of condition is expected for task relevance because both tutorials present the same task demonstrations. Self-efficacy, however, is likely to be positively affected by the reviews. As reviews summarize the main steps for task accomplishment, they may increase the participant's confidence about (future) task completion.

Research question 3: *How well do the video tutorials support task performance and learning, and is there an effect of condition?*

The video tutorials are expected to be equally effective in supporting task performance during training, if only because participants can always consult the demonstration videos. After training, video access is blocked when learning is assessed. Participants are tested immediately after training and again one week later. Because the review videos support retention, they should yield higher scores on both an immediate post-test and a delayed post-test taken one week after training.

Method

Participants

The participants consisted of 55 students (mean age 13 years; range 12.0–14.4) at two middle schools in Germany. Accordingly, all study materials were in German. The 33 male students and 22 female students were from two seventh grade classrooms. Individuals were randomly assigned to condition, after stratification for classroom.

Instructional materials

The *video tutorials* teach several formatting tasks in the German version of Microsoft Word 2007. The tasks are organized into three 'chapters'. Chapter 1 demonstrates how to adjust the left and right margins for an entire document, in two task videos. Chapter 2 models the formatting of paragraphs, citations and lists, in four videos. Chapter 3 demonstrates how Word can create a table of contents, in five videos. The tasks all revolve around improving the lay-out of reports that the participants must regularly produce for school. The tasks are presented in a simple-to-complex sequence.The tutorials are presented on a website that is divided into two areas (Figure 1). The area on the left presents a clickable table of contents. The table of contents is always visible to afford easy and permanent access to the recorded demonstrations. Chapter titles (purple background) organize similar tasks. Task titles (light blue background) link to the videos, as signalled by an icon at the end.

After a participant clicks on a task title, the light blue background colour changes to orange and the demonstration video comes up on the right-hand side of the website. In addition, a transparent toolbar automatically appears at the bottom, allowing the participant to start the video playing, pause, resume and stop (Figure 1). The toolbar can also be used to increase or decrease the sound level. A progress bar shows how far the video has progressed.

Demonstration videos contain information about the goal, the required participant actions and the effects of these actions. The narrator, a native male speaker of German, begins by introducing the upcoming task. The participants are told about the nature of a formatting problem ('This text has margins that are too narrow') and the solution. This section of the narrative should have a positive effect on the participants' motivation, contributing to goal setting (Grossman et al., 2013; van der Meij & van der Meij, 2013). After that, the narrator consistently tells participants about required actions involving the input device (e.g., 'Click the left mouse button') and the interface (e.g., 'Drag the margin to 2.5 centimeters'). The effects of these actions on the

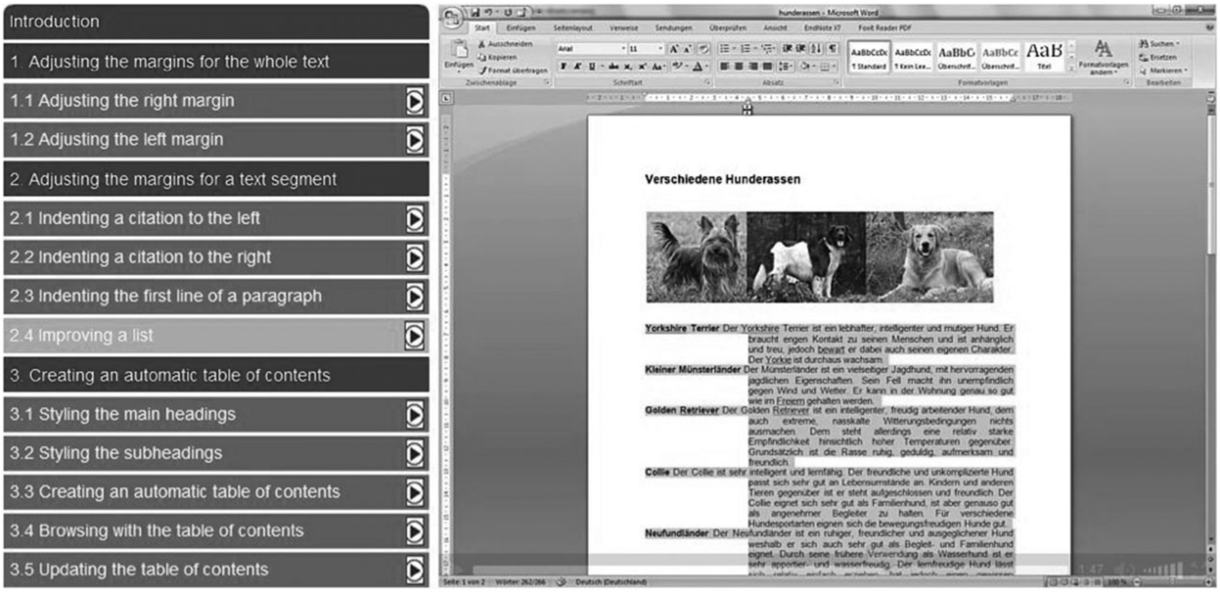


Figure 1 Screen Shot of the Website (Translated Version)

interface are also shown, and the narrator regularly draws the participants' attention to these changes, using standard phrases such as, 'A window appears with the text …' and 'You now see …' Zooming-in and highlighting (Figure 2) occasionally complement these comments, in order to further emphasize pertinent screen objects or areas. Finally, after the task demonstration has been completed there is a brief pause of 2 s. In all, there were 11 task demonstration videos, with a mean length of 1.14 min (range 0.48–1.46).

Review videos summarize task demonstrations. They appear automatically a few seconds after a demonstration has finished. All reviews begin with the announcement 'You're finished now, but remember…' Thereafter, the narrator frames his comments in the 'I'-form. This indicates the difference between the review and the task demonstration and minimizes the need for participants to recode statements into personal action plans (e.g., 'First, I must click on […] Then, I should select […] Finally, I press the TAB-key'). The review video concentrates on the (sub)goals and the actions that need to be performed. Compared with the demonstrations much less attention is given to the system reactions. Just as in the demonstrations, the reviews include animated screen displays with signalling. Reviews take from 13 to 26 s. The mean length of a task demonstration video with review is 1.31 min (range 1.02–2.06).

During training, participants are instructed to use *practice files* that have been created especially for these tasks. These files are accessible from a folder with the student's name that is on the computer desktop. Practice files minimize the need for task-irrelevant actions, such as typing, and they include few distracting formatting features (van der Meij & Carroll, 1998). In addition, these files standardize practice; they make task completion efforts comparable across conditions. The *training time measure* is based on saved modified practice files, as is *task performance success* during training. Inspection of the practice files later revealed that participants had forgotten to save about 10% of these task files. In data analyses these omissions were counted as incorrect solutions.

A paper *instruction booklet* provides participants with a task scenario and supports them in switching between viewing the video and engaging in practice. The booklet sets out a task sequence that is identical with that of the table of contents on the website. This task sequence guides participants to work though the materials in a fixed order. Chapter titles organize the task titles. Under each task title, the booklet instructs the participant first to watch the video, and then to engage in practice (Figure 3). The booklet also includes a few (repeated) questions to assess motivational mediators (i.e,, mood and flow). The findings for these mediators favoured the Review condition, but are not reported here.

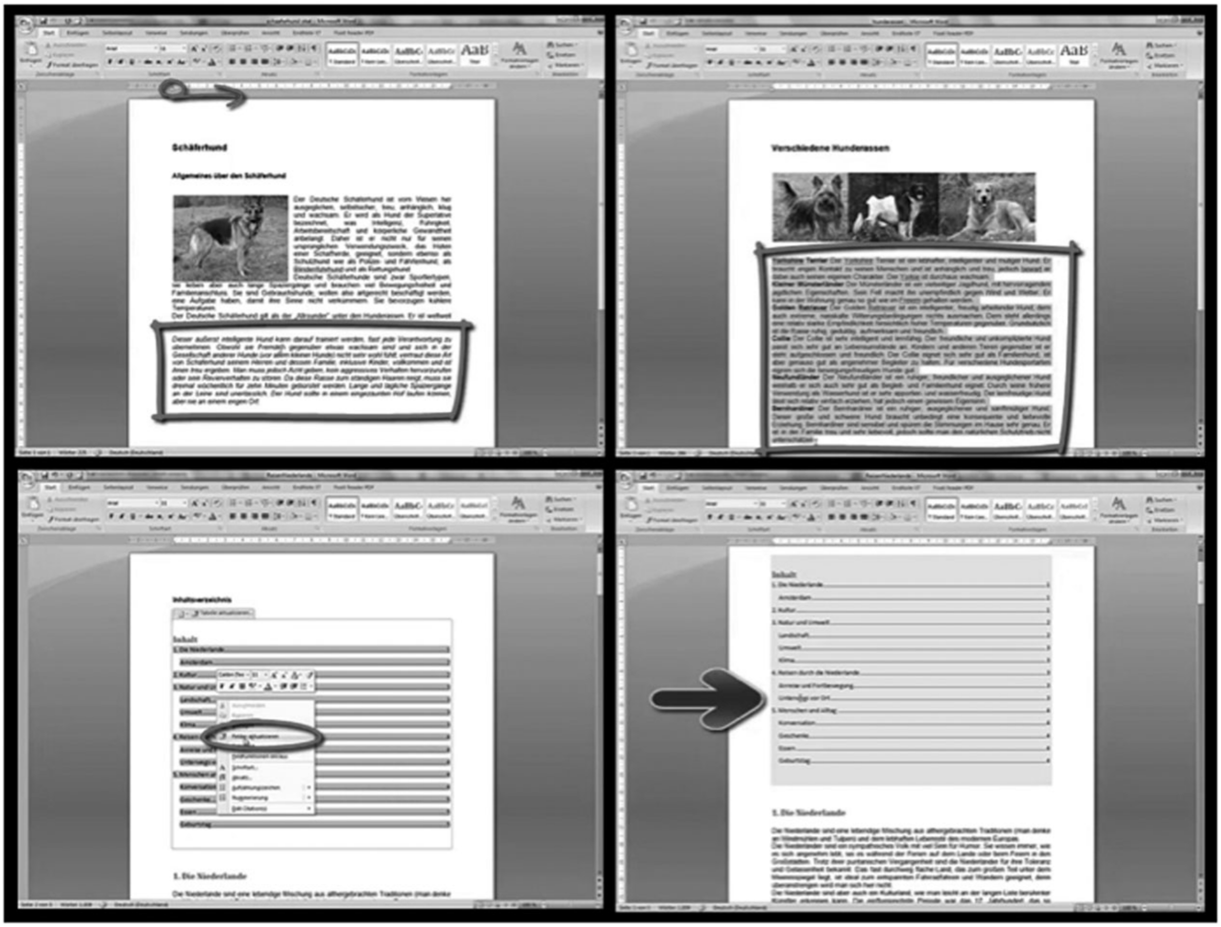


Figure 2 Examples of the Signaling Used in the Videos

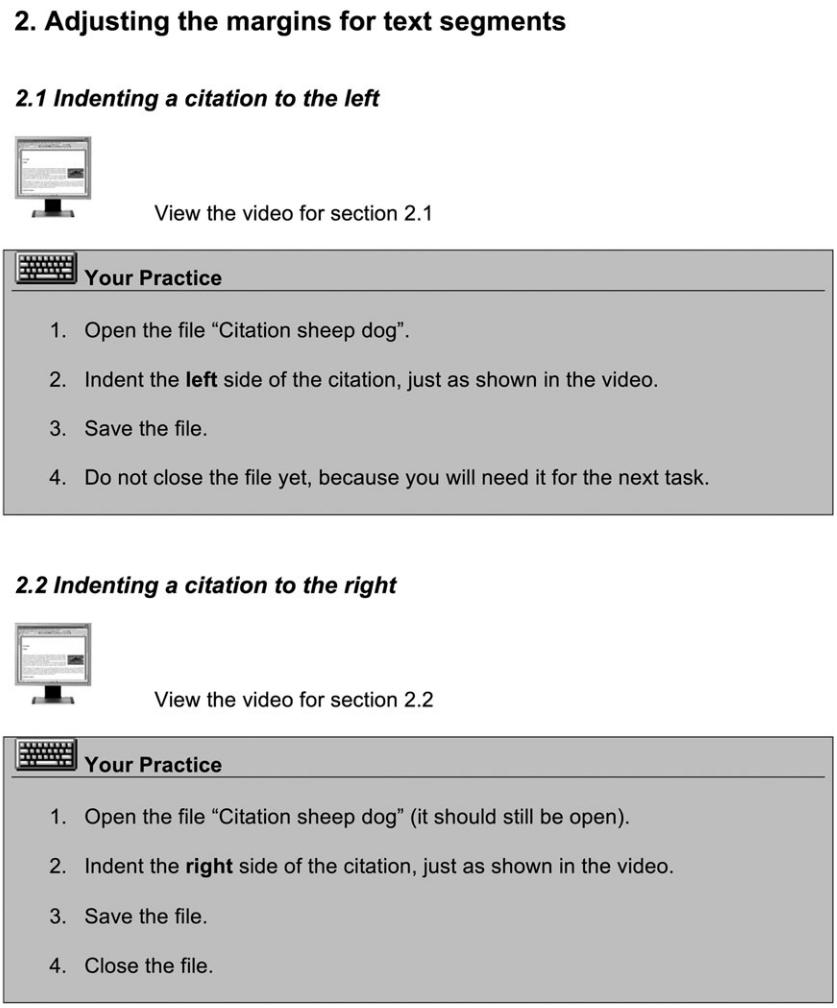
Instruments

An *Initial Experience and Motivation Questionnaire* (*IEMQ*) measures the participants' experience and motivation before training to check on their random distribution for these across conditions and to provide a baseline measure. The IEMQ presents a screenshot for each of the six training tasks and asks three questions about that task: Do you ever face this situation? (experience), How often do you need to complete this task? (task relevance), and How well do you think you can complete this task? (self-efficacy). In all, the IEMQ holds six questions about task relevance and six questions about self-efficacy. Answers are given on a 7-point Likert scale which range from never (1) to always (7), or very poorly (1) to very well (7). Reliability analyses with Cronbach's alpha indicated satisfactory results for the three measures (i.e., experience $\alpha = 0.72$; task relevance $\alpha = 0.83$; self-efficacy $\alpha = 0.80$).

The Final Motivation Questionnaire asks participants to rate the relevance of the trained tasks and to report their self-efficacy for achieving these tasks in future. There are eight items about task relevance, for example, I find it important to present nice lists and I find it important to present a good table of contents. Likewise, there are eight items about self-efficacy, for example, I can now indent the first line of a paragraph and I now know how to automatically create a table of contents. Answers are given on a 7-point Likert-scale. The reliability score, Cronbach's alpha, for task relevance was 0.87. For self-efficacy it was 0.86.

Knowledge tests

A *pre-test* asks participants to demonstrate their formatting skills by completing the type of tasks that are to be trained. During this test the participants are asked to



2. Adjusting the margins for text segments

2.1 Indenting a citation to the left

View the video for section 2.1

Your Practice

1. Open the file "Citation sheep dog".
2. Indent the **left** side of the citation, just as shown in the video.
3. Save the file.
4. Do not close the file yet, because you will need it for the next task.

2.2 Indenting a citation to the right

View the video for section 2.2

Your Practice

1. Open the file "Citation sheep dog" (it should still be open).
2. Indent the **right** side of the citation, just as shown in the video.
3. Save the file.
4. Close the file.

Figure 3 A Section from the Instruction Booklet (Translated Version)

modify the format of pre-test files. An *immediate post-test* asks participants to complete the same formatting tasks addressed in the training, using post-test files that differ only in appearance from pre-test or practice files. The *delayed post-test* is similarly constructed. Correct task completion is worth 1 point and an incorrect attempt is worth 0 points. All performance test measures are timed (maximum 20 minutes). Test scores are converted to a percentage of possible points.

Procedure

The experiment was conducted in two sessions that took place in the schools' computer laboratories. Each computer was labelled with a number and equipped with the files and instruments for the study. In addition, there were headphones for the participants to use during training. The first session began with a 5-min introduction that informed participants about the Word training they would receive. After that, the IEQM and pre-test were completed (maximum 20 min). After a short break, this was followed by another (10 min) introduction, in which the training procedure and the use of the instruction booklet were explained. Participants could also practice site navigation and file handling with a scaled-down version of the website. Furthermore, participants were instructed to wear headphones during training, to work individually and to ask for help only when they experienced technical problems. During training, participants could always consult the videos. The maximum training time was 40 min. After that, participants completed the *Final Motivation Questionnaire* and took the post-test

(20 min). During testing, participants were not allowed to consult the videos. The second session took place seven days later. In this session, participants took the delayed post-test (20 min).

Analysis

A check on random distribution of participants across conditions revealed no statistically significant differences for age, $F(1,54) = 1.03$, n.s., or gender $\chi^2(1,55) = 0.045$, n.s. Conditions also did not differ on IEMQ-scores. Repeated measures ANOVAs were computed to gauge changes over time within conditions. ANCOVAs were computed to examine the effect of condition on motivation, task performance and learning using the IEMQ or pre-test score as covariate. Tests on the assumption of homogeneity of variance indicated no violations. Likewise, there was no violation of the assumption of homogeneity of regression slopes in the ANCOVAs. For some measures the degrees of freedom varied slightly, because of missing data. One outlier was removed for the motivation measures after training. The tables present only the data from students with complete data sets. Tests were one-tailed for directional predictions (indicated with the test result) and two-tailed for all other cases, with alpha set at 0.05. Cohen (1988) d-statistic is used to report effect size. These tend to be qualified as small for $d = 0.2$, medium for $d = 0.5$ and large for $d = 0.8$.

Results

Training time

Participants in the Control condition needed an average of 31 min ($SD = 2.42$) to complete training. In the Review condition the average *training time* was 29.3 min ($SD = 2.65$). Contrary to the prediction, there was a statistically significant difference in training time between conditions, $F(1,54) = 5.81$, $p = 0.019$, $d = 0.67$. Participants in the Review condition completed training faster.

Motivation before and after training

The scores for *task relevance* before training indicate the presence of a relatively low level of prior task interest (Table 1). After training the task relevance appraisals were significantly higher, $F(1,50) = 165.33$, $p < 0.001$, $d = 2.51$. The resulting scores for task relevance were substantially above the scale midpoint. An analysis of covariance on task relevance-after, with task relevance-before as a covariate, showed no effect of condition, $F(1,51) = 1.53$, n.s., which was the predicted finding.

The initial scores for *self-efficacy* were around the scale midpoint, which suggests that participants began training with some degree of confidence in their capacities to deal with the training tasks (Table 1). After training self-efficacy belief was significantly higher, $F(1,50) = 61.68$, $p < 0.001$, $d = 1.57$, yielding scores that were substantially above the scale midpoint. An analysis of covariance for self-efficacy-after, with self-efficacy-before as a covariate, showed a significant effect of condition, $F(1,51) = 4.29$, $p = 0.022$ (one-sided). As predicted, participants in the Review condition showed greater increase in self-efficacy ratings after training.

Task performance and learning

Table 2 displays the findings for the effectiveness of the video tutorial as a support for task performance during training. The table shows high absolute levels of performance success. As predicted, there was no difference between conditions, $F < 1$. High as they are, the outcomes are still somewhat conservative because we treated missing data as incorrect solutions. Had we excluded those

Table 1. Means (Standard Deviations) for Task Relevance[a] and Self-efficacy[a] by Condition

	Task relevance				Self-efficacy			
	Before		After		Before		After	
Condition	*M*	(*SD*)	*M*	(*SD*)	*M*	(*SD*)	*M*	(*SD*)
Review ($n = 28$)	2.70	(1.35)	6.07	(0.66)	4.10	(1.70)	5.97	(0.52)
Control ($n = 24$)	3.24	(1.65)	5.80	(0.94)	3.73	(1.57)	5.61	(0.67)
Total ($n = 52$) $^{+}$	*2.95*	(*1.50*)	*5.94*	(*0.80*)	*3.93*	(*1.64*)	*5.81*	(*0.62*)

[a]Scale maximum is 7. A higher score indicates higher appreciation.

Table 2. Mean Success Rates (Standard Deviations) for Pre-test, Training, Immediate Post-test and Delayed Post-test by Condition

	Pre-test		Training		Immediate Post-test		Delayed Post-test	
Condition	*M*	(*SD*)	*M*	(*SD*)	*M*	(*SD*)	*M*	(*SD*)
Review ($n = 28$)	23.7	(15.0)	88.4	(15.6)	86.2	(18.1)	89.3	(13.5)
Control ($n = 24$)	22.9	(18.3)	86.5	(19.1)	77.1	(19.7)	81.8	(23.3)
Total ($n = 52$)	*23.3*	(*16.4*)	*87.5*	(*17.1*)	*82.0*	(*19.2*)	*85.8*	(*18.9*)

data, a mean success score of 98.3% for training tasks (in both conditions) would have been obtained.

The findings for the effectiveness of the video tutorials as a support for learning are likewise positive. An average success rate of 82.1% was achieved on the immediate post-test (Table 2). Compared with the mean pre-test score of 23%, the increase was both statistically significant and substantial, F (1, 53) = 331.97, $p < 0.001$, $d = 3.34$. A slightly higher success rate was even seen on the delayed post-test. Compared with the mean pre-test score, this increase was also both statistically significant and substantial, F (1, 50) = 336.88, $p < 0.001$, $d = 3.57$.

An analysis of covariance on the immediate post-test, with pre-test scores as covariate, yielded a statistically significant effect for condition, F (1,54) = 2.90, $p = 0.048$ (one-sided) in favour of the Review condition, as predicted. However, contrary to prediction, the same analysis showed no difference between conditions for the delayed post-test, F (1,51) = 2.04, n.s.

Discussion and conclusion

Contrary to expectations, a significant effect in training time was found in favour of the Review condition. Perhaps, the time difference reflects the benefit of having a short recap before practice. That is, participants who have just seen a review are likely to have better retention of the main steps in a procedure and would therefore need to check back less often to support task execution during practice.

The video tutorials had a strong and positive effect on motivation. Measures of the participants' task relevance indicated that training significantly increased this perception. A similar finding was obtained for self-efficacy, indicating that participants both found the formatting tasks meaningful and felt confident that they could deal with such tasks in the future. The data also revealed high post-training results for motivation. After training, the participants rated task relevance and self-efficacy at over 80% of the scale maximum. For task relevance this finding is even more remarkable, as participants started out with scores initially below the scale mid-point.

As expected, a significant difference between conditions was found for self-efficacy, favouring the Review condition. Presumably, this is an effect of reminding the user what it takes to achieve task completion. That is, when a review recapitulates the key steps and actions in a task completion process it conveys the impression that the task is manageable, requiring only a few actions to round it off successfully.

The video tutorials also significantly and substantially improved success on task performance. From an initial success rate of 23%, the scores increased to a success rate of 87% on the training tasks. Because participants could consult the videos during training, this score signals the effectiveness of the tutorials as a job-performance aid (van der Meij et al., 2009). In addition, significant and substantial learning effects were found. Test scores for the immediate and delayed post-test were 82% and 85%, respectively. In other words, the absolute level of participants' task and test performance was high, with and without review.

The prediction that the Review condition would result in greater learning was partly supported. A significant effect of condition favouring the Review condition was found, but only on the immediate post-test. We ascribe this effect to the retention process that the reviews set out to support. In retention, users must transform their observations of the demonstration videos into symbolic codes that are stored for future behaviour. The reviews address this retention process, as they present the steps towards task completion in condensed format. They show what a mental replay or cognitive rehearsal of the steps involved in task completion would look like.

While a positive effect of the review was found for the immediate post-test, there was no difference between conditions for the delayed post-test. The review is a concise recap of the demonstrated procedure. It bears a great visual resemblance to the training, as the review included the same Word file and the same signalling techniques. These similarities may have given participants the impression that the review was more or less redundant which could reduce attention.

To avoid such a similarity effect, future studies might want to tweak the design of the review a bit, making it slightly more different from the demonstration. Fundamental research on the desirable difference hypothesis suggests how important subtle design variations can be for retention and recall. According to this hypothesis, multimedia presentations benefit from on-screen texts that represent the narrative in abridged form (Bjork & Bjork, 2011; Yue, Bjork, & Bjork, 2013). By creating a small discrepancy between what the narrator says and what is written on the screen, the participant is stimulated to pay close attention to both sources and process the information at a deeper level. Following a similar line of reasoning, the review can probably be made stronger by reducing its similarity to the demonstration. The review should avoid giving the impression that it is a (concise) repetition, which can be achieved by using different visuals. That is, the review should present a different Word file and use different signalling techniques than the demonstration. Will this make the review stronger and affect retention and recall as it does in the research on desirable differences? Only empirical research can tell whether such variations have a positive effect on retention, or whether they tax participants too much.

A limitation of the research is that the training context is not the ordinary self-study situation. The participants were seated in their own classroom, and they were expected to process the videos and complete their training tasks for the duration of the experiment. Another limitation is that no insights were gathered about usage patterns for the videos. Did the participants first view a task video in full and then attempted to achieve that task on the training file, or did they start earlier, using the demonstration step by step for guiding their actions? Recording user logs would give insight into such usage patterns. With such logs it is also possible to discover the sections in a video that call for a revision. A signal of such a weak point would be a moment in which many participants temporarily stop the video. These limitations notwithstanding, the video tutorials in the study did yield substantial engagement and effectiveness as called for in the usability criteria set forth by the ISO-9241 standard.

Acknowledgement

We would like to thank Philip Neuß for his help in conducting the experiment described in this paper.

References

Akinlofa, O. R., O'Brian Holt, P., & Elyan, E. (2013). Domain expertise and the effectiveness of dynamic simulator interfaces in the acquisition of procedural motor skills. *British Journal of Educational Technology*, *44*, 810–820. doi:10.1111/j.1467-8535.2012.01364.x.

Amadieu, F., Mariné, C., & Laimay, C. (2011). The attention-guiding effect and cognitive load in the comprehension of animations. *Computers in Human Behavior*, *27*, 36–40. doi:10.1016/j.chb.2010.05.009.

Ayres, P., Marcus, N., Chan, C., & Qian, N. (2009). Learning hand manipulative tasks: When instructional animations are superior to equivalent static representations. *Computers in Human Behavior*, *25*, 348–353. doi:10.1016/j.chb.2008.12.013.

Bandura, A. (1986). *Social foundations of thought and actions: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.

Bandura, A. (1997). *Self-efficacy. The exercise of control*. New York, NY: Freeman and Company.

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64): New York, NY:, Worth Publishers.

Brucker, B., Scheiter, K., & Gerjets, P. (2014). Learning with dynamic and static visualizations: Realistic details only benefit learners with high visuospatial abilities. *Computers in Human Behavior*, *36*, 330–339. doi:10.1016/j.chb.2014.03.077.

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

Ginns, P., Martin, A. J., & Marsh, H. W. (2013). Designing instructional text in a conversational style: A meta-analysis. *Educational Psychology Review*, *25*, 445–472. doi:10.1007/s10648-013-9228-0.

Grossman, R., Salas, E., Pavlas, D., & Rosen, M. A. (2013). Using instructional features to enhance demonstration-based training in management education. *Academy of Management Learning and Education*, *12*, 219–243. doi:10.5465/amle.2011.0527.

Guo, P. J., Kim, J., & Rubin, R. (2014). How video production affects student engagement: An empirical study of MOOC videos. Paper presented at the L@S '14, Atlanta, GA.

Hartley, J., & Davies, I. K. (1976). Preinstructional strategies: The role of pretests, behavioral objectives, overviews and advance organizers. *Review of Educational Research*, *46*, 239–265. doi:10.2307/1170040.

Hartley, J., & Trueman, M. (1982). The effects of summaries on the recall of information from prose: Five experimental studies. *Human Learning*, *1*, 63–82.

Hartley, J., Goldie, M., & Steen, L. (1976). The role and position of summaries: Some issues and data. *Educational Review*, *31*, 59–65. doi:10.1080/0013191790310107.

Hoogerheide, V., Loyens, S. M. M., & van Gog, T. (2014). Comparing the effects of worked examples and modeling examples on learning. *Computers in Human Behavior*, *41*, 80–91. doi:10.1016/j.chb.2014.09.013.

Jin, S.-H. (2013). Visual design guidelines for improving learning from dynamic and interactive digital text. *Computers & Education*, *63*, 248–258. doi:10.1016/j.compedu.2012.12.010.

Kartal, G. (2010). Does language matter in multimedia learning? Personalization principle revisited. *Journal of Educational Psychology*, *102*, 615–624. doi:10.1037/a0019345.

Koumi, J. (2013). Pedagogic design guidelines for multimedia materials: A call for collaboration between practitioners and researchers. *Journal of Visual Literacy*, *32*(2), 85–114.

Lemarié, J., Lorch, R. F., Eyrolle, H., & Virbel, J. (2008). SARA: A text-based and reader-based theory of signaling. *Educational Psychologist*, *43*, 27–48. doi:10.1080/00461520701756321.

Lloyd, S. A., & Robertson, C. L. (2012). Screencast tutorials enhance student learning of statistics. *Teaching of Psychology*, *39*(1), 67–71. doi:10.1177/0098628311430640.

Lowe, R., Schnotz, W., & Rasch, T. (2011). Aligning affordances of graphics with learning task requirements. *Applied Cognitive Psychology*, *25*, 425–459. doi:10.1002/acp.1712.

McLaughlin Cook, N. (1981). Summaries: Further issues and data. *Educational Review*, *33*(3), 215–222. doi:10.1080/0013191810330305.

van der Meij, H., & Carroll, J. M. (1998). Principles and heuristics for designing minimalist instruction. In J. M. Carroll (Ed.), *Minimalism beyond the Nurnberg funnel* Cambridge. MA: MIT Press.

van der Meij, H., & van der Meij, J. (2013). Eight guidelines for the design of instructional videos for software training. *Technical Communication*, *60*, 205–228.

van der Meij, H., & van der Meij, J. (2014). A comparison of paper-based and video tutorials for software learning. *Computers & Education*, *78*, 150–159.

van der Meij, J., & van der Meij, H. (2015). A test on the design of a video tutorial for software training. *Journal of Computer-Assisted-Learning*, *31*, 116–132.

van der Meij, H., Karreman, J., & Steehouder, M. (2009). Three decades of research and professional practice on software tutorials for novices. *Technical Communication*, *56*, 265–292.

Merkt, M., Weigand, S., Heier, A., & Schwan, S. (2011). Learning with videos vs. learning with print: The role of interactive features. *Learning and Instruction*, *21*, 687–704. doi:10.1016/j.learninstruc.2011.03.004.

van Merriënboer, J. J. G., Kirschner, P. A., & Kester, L. (2003). Taking the load off a learner's mind: Instructional design for complex learning. *Educational Psychologist*, *38*, 5–13. doi:10.1207/S15326985EP3801_2.

Pintrich, P. R., & Schunk, D. H. (2002). *Motivation in education. Theory, research, and applications* (2nd ed.). Upper Saddle River, NJ: Merrill Prentice Hall.

Plaisant, C., & Shneiderman, B. (2005). Show me! Guidelines for recorded demonstration. Paper presented at the 2005 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC'05), Dallas, Texas. http://www.cs.umd.edu/localphp/hcil/tech-reports-search.php?number=2005-02

Ploetzner, R., & Lowe, R. (2012). A systematic characterisation of expository animations. *Computers in Human Behavior*, *28*, 781–794. doi:10.1016/j.chb.2011.12.001.

Reichelt, M., Kämmerer, F., Niegemann, H. M., & Zander, S. (2014). Talk to me personally: Personalization of language style in computer-based learning. *Computers in Human Behavior*, *35*, 199–210. doi:10.1016/j.chb.2014.03.005.

Reisslein, J., Atkinson, R. K., Seeling, P., & Reisslein, M. (2006). Encountering the expertise reversal effect with a computer-based environment on electrical circuit analyses. *Learning and Instruction*, *16*, 92–103. doi:10.1016/j.learninstruc.2006.02.008.

Rosen, M. A., Salas, E., Pavlas, D., Jensen, R., Fu, D., & Lampton, D. (2010). Demonstration-based training: A review of instructional features. *Human Factors*, *52*, 596–609. doi:10.1177/0018720810381071.

Schwan, S., & Riempp, R. (2004). The cognitive benefit of interactive videos: Learning to tie nautical knots. *Learning and Instruction*, *14*, 293–305. doi:10.1016/j.learninstruc.2004.06.005.

Smith, P. L., & Ragan, T. J. (2005). *Instructional design* (3rd ed.). Hoboken, NJ: Wiley.

Spanjers, I. A. E., van Gog, T., Wouters, P., & van Merrienboer, J. J. G. (2012). Explaining the segmentation effect in learning from animations: The role of pausing and temporal cueing. *Computers & Education*, *59*, 274–280. doi:10.1016/j.compedu.2011.12.024.

Stiller, K. D., Freitag, A., Zinnbauer, P., & Freitag, C. (2009). How pacing of multimedia instructions can influence modality effects: A case of superiority of visual texts. *Australasian Journal of Educational Technology & Society*, *25*, 184–203.

Tversky, B., Bauer-Morrison, J., & Bétrancourt, M. (2002). Animation: Can it facilitate? *International Journal of*

Human-Computer Studies, *57*, 247–262. doi:10.1006/ijhc.2002.1017.

Wistia. (2012). Does length matter? Retrieved from http://wistia.com/blog/does-length-matter-it-does-for-video-2k12-edition

Witteman, M. J., & Segers, E. (2010). The modality effect tested in children in a user-paced multimedia environment. *Journal of Computer Assisted Learning*, *26*, 132–142. doi:10.1111/j.1365-2729.2009.00335.x.

Wouters, P., Paas, F., & van Merriënboer, J. J. G. (2010). Observational learning from animated models: Effects of studying-practicing alternation and illusion of control on transfer. *Instructional Science*, *38*, 89–104. doi:10.1007/s1121-008-9079-0.

Yue, C. L., Bjork, E. L., & Bjork, R. A. (2013). Reducing verbal redundancy in multimedia learning: An undesired desirable difficulty? *Journal of Educational Psychology*, *105*, 266–277. doi:10.1037/a0031971.