Computers & Education 114 (2017) 164–174

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Reviews in instructional video

Hans van der Meij

University of Twente, Faculty of Behavioural, Management and Social Sciences, Department of Instructional Technology, P.O. Box 217, 7500 AE Enschede, The Netherlands

A R T I C L E I N F O

Article history:
Received 19 December 2016
Received in revised form 21 June 2017
Accepted 3 July 2017
Available online 4 July 2017

Keywords:
Software tutorial
Video
Reviews
Demonstration-based training
Multimedia learning theory

A B S T R A C T

This study investigates the effectiveness of a video tutorial for software training whose construction was based on a combination of insights from multimedia learning and Demonstration-Based Training. In the videos, a model of task performance was enhanced with instructional features that were intended to be particularly effective insofar as they addressed four key processes in observational learning (i.e., attention, retention, reproduction and motivation). An experiment with two conditions was reported. The control condition consisted of only demonstration videos. The experimental condition included a review after task demonstration to provide additional support for retention. The videos taught Word formatting tasks. The 73 participants came from elementary and secondary school. During training, video playing was followed by task practice. After training, a post-test was administered. Engagement data showed that demonstration videos were played almost completely (93%). Reviews fared worse (32%). Motivation increased significantly with training regardless of condition. Task performance also increased significantly from pre-test (29%) to training (84%) and post-test (71%). In addition, results for performance during and after training were significantly better for the experimental condition than the control condition. The discussion argues that the demonstration videos provide a viable way to support task completion. To further improve learning, better understanding of learners' retention processes is needed.

1. Introduction

Videos have rapidly become immensely popular. Technological developments have contributed to their growth in popularity, with computers becoming substantially cheaper, faster and equipped with storage capacities that have expanded from megabytes to gigabytes and terabytes. Likewise, programs for producing videos have become cheaper and easier to use. Improved possibilities for sharing have further contributed to the rising popularity of videos. Ready distribution is now commonplace, with YouTube remaining the top source for all types of videos.

This paper's focus is on how-to videos. Their primary function is to support task performance; they are job aids whose foremost purpose is to enable or guide task completion (Kim et al., 2014). Examples of how-to videos that support task completion can be found on the websites of manufacturing companies, among other places. Ikea is one firm that has produced such videos. Ikea videos can be found in a "How to build" section on the company website, but the same videos are more conveniently located on YouTube (Ikea, 2017; May 1). All Ikea assembly videos on YouTube have the same look and feel (and background music). A brief introduction shows the materials, tools and people (1 or 2) needed for assembly. Step-by-step

E-mail address: H.vanderMeij@utwente.nl.

http://dx.doi.org/10.1016/j.compedu.2017.07.002
0360-1315/

instructions follow, in which real people model correct task completion. The actors, dressed in Ikea uniform, silently demonstrate what to do; there is no spoken narrative. Their modeling actions are supported with brief captioned descriptions (e.g., "Make sure grooved sides are facing in"), signaling techniques (arrows, zooming in and out) and small, on-screen displays of the required (or wrong) materials. The statistics for the YouTube video on, say, the assembly of the PAX wardrobe, reveal that it has been viewed 316,200 times, which illustrates its popularity. Other Ikea assembly videos had reached between 59,175 and 447,961 views when last checked.

Some how-to videos must also serve an additional role, however; namely, to support learning. In addition to enabling task completion, these videos should support procedural knowledge development. In these cases, the videos are tutorials that must establish a basic level of task competence that equips users with the capacity to (quickly learn to) perform other, similar tasks (e.g., Carroll & Rosson, 1987; Fu & Gray, 2004). Videos for software training often fall within this category.

Production of such videos can be a profitable business, as shown by the example of Lynda, an online education company. In 2015, LinkedIn purchased the website Lynda.com for the amount of $1.5 billion. One reason for the deal was so that the LinkedIn website could indicate an available job plus skills requirements, and direct a possible applicant to a video tutorial from Lynda.com that trained the desired skills (Kosoff, 2017; May 2). Software makers also produce video tutorials for their users. For instance, Adobe offers several video tutorials on Premiere Pro (Adobe, 2017; May 12), Microsoft does the same for Word (Microsoft, 2017; May 12) as does TechSmith for Camtasia (TechSmith, 2017; May 10).

On the Lynda, Adobe, Microsoft and TechSmith websites, tutorials for software training are framed as recorded demonstrations. They display an animated screen recording that is accompanied by a spoken narrative. This is considered the standard and preferred format for video instruction for software usage (Plaisant & Shneiderman, 2005). Unfortunately, there is no published information about the design characteristics and effectiveness of the tutorials on these company websites, to our knowledge. The question thus arises what are the design characteristics of an effective software tutorial. This paper suggests that a primary vehicle for facilitating immediate task performance in such a tutorial is a demonstration. Because software tutorials are a special type of how-to video that must also facilitate learning, additional instructional support is needed. This study investigates whether complementary video reviews enhance retention over and above other instructional features.

2. A theoretical model for software training with video

Combined insights from Demonstration-Based Training (DBT) and multimedia learning theory form the basis for our construction of a video tutorial for software training. DBT claims that a model of task performance that the user can mimic should form the heart of a DBT-based video tutorial for software training. It further states that instructional support is needed to increase learning from such a model (Grossman, Salas, Pavlas, & Rosen, 2013; Rosen et al., 2010). According to DBT, these supportive features should address the main processes involved in observational learning that were distinguished by Bandura (1986), namely, attention, retention, practice and motivation. Multimedia learning theory (see Mayer, 2014a) complements the DBT view with important insights on (limitations of) cognitive processing and with empirically tested design principles. The influence of both views is seen in the theoretical model for video construction that formed the basis for the videos that were constructed and tested in the present study (see Fig. 1). A recent paper by Brar and van der Meij (2017) gives a detailed account of the model. The interested reader is referred to that paper for a description.

Notice that the theoretical model includes both user characteristics and situational variables. One user characteristic that should always be considered when seeking to optimize observational learning is prior knowledge. Multimedia research generally shows that low prior knowledge users benefit considerably from instructional support, while high prior knowledge users may not need such support, and can even be hindered by its presence (Kalyuga, 2007; Mayer, 2014a). Situational variables constrain the design guidelines that can be applied. For example, the theoretical model presented in Fig. 1 concentrates on the individual user working alone at his or her computer. We have therefore excluded instructional features that involve instructor-led support.

In the theoretical model, the instructional features are connected to the observational learning process on which they presumably have the strongest impact, but clearly they can affect more than a single process. To support learning, a minimum requirement would be that at least one instructional feature should be included for each of the four observational learning processes. But in a more realistic, practical scenario, several features for each process should be incorporated in the design.

3. On including reviews in instructional videos for software training

Several recent studies (H. van der Meij, 2014; H. van der Meij & van der Meij, 2014; J. van der Meij & van der Meij, 2015) have investigated the effectiveness of video tutorials for software training that were designed according to the theoretical model presented in Fig. 1. These studies all involved very similar demonstration videos for software training, with regard to both the content they taught (i.e., formatting tasks in Word), and the presence of certain instructional features coupled with the demonstration of task performance. Mean success rates of over 80% for (aided) task performance during training were consistently reported. In other words, the data revealed that users understood the videos well enough to let them perform the modeled procedures. However, when learning of the instructed tasks was tested, a significant decline was found, with mean success rates often dropping to below 70% for (unaided) task performance. This finding points to a lack of retention-related

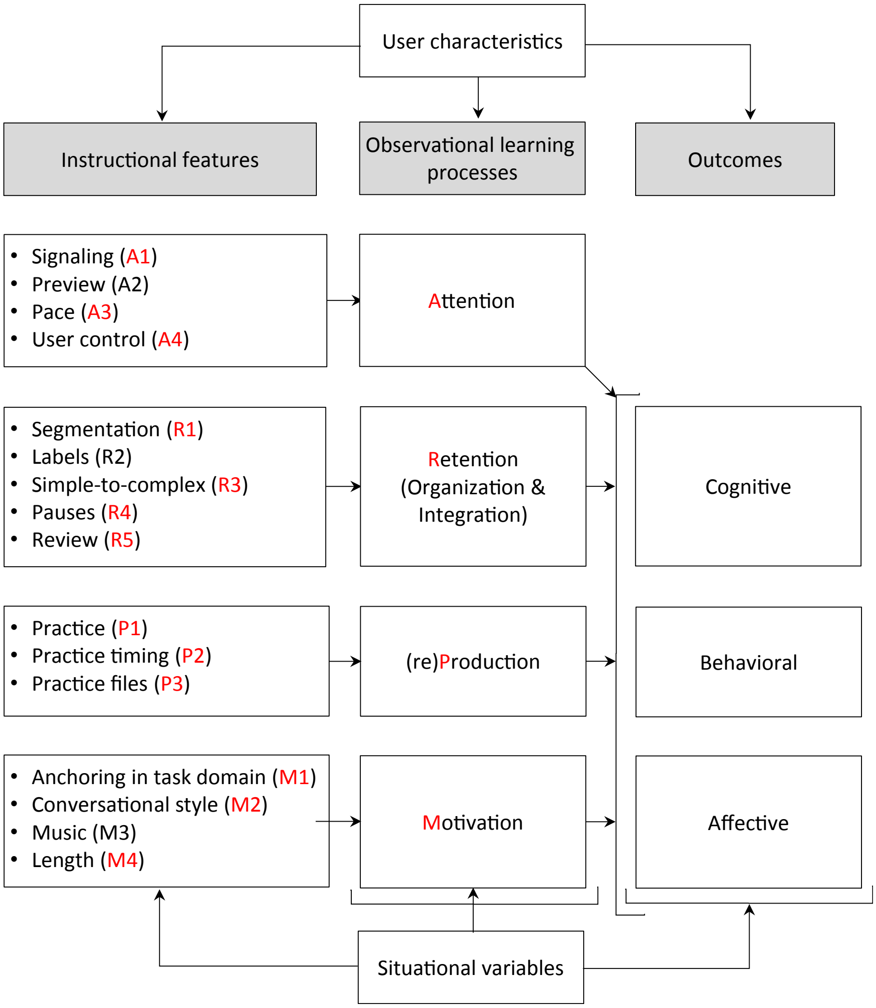

Fig. 1. Theoretical Model for video construction for software training. Highlighted features (in red) were incorporated in the design of the tutorial in this study. These features and their supporting evidence are described in the Method section. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

processing. In other words, these studies suggest that additional, dedicated instructional support may be needed to enhance storing a procedure in long-term memory.

Reviews would appear to be ideal for supporting this aspect of the user's retention process. As a recap of the main events involved in task performance, in order, the review provides the user with an overview of the main issues or events in a procedure, and gives a summary of how task completion is organized. A review also provides the user with a concise repetition or replay that should benefit memory storage. Alternatively, its presence may offer the user a second chance to learn; it may compensate for any possible mind wandering when watching the demonstration. In addition, a review can serve as a frame of reference for the user's own summary of a procedure. When a demonstration video is complemented with a review, the user can compare it to the self-constructed summary. When the user notices an important discrepancy, this can result in replaying a section of the demonstration video, among other things.

Very little research appears to have been done on reviews. A literature search for empirical studies on the effectiveness of reviews in instructional videos revealed no hits. In addition, the meta-analysis of expository animations by Ploetzner and Lowe (2012) did not report a single case involving reviews among the 44 empirical studies that were analyzed. When our search was extended to include summaries used with texts, the pursuit was only slightly more fruitful, yielding only a few older studies (e.g., Christensen & Stordahl, 1955; Hartley, Goldie, & Steen, 1976; McLaughlin Cook, 1981; Vezin, Berge, & Mavrelis, 1973). Recent research appears to have ignored predefined text summaries, concentrating instead on summaries self-generated by students (e.g., Gil, Bråten, Vidal-Abarca, & Strømsø, 2010; Leopold, Sumfleth, & Leutner, 2013; Westby, Culatta, Lawrence, & Hall-Keyon, 2010). The older studies were discussed by Hartley and Trueman (1982), who also conducted five consecutive experiments on how summaries and their placement affect retention. The overall finding for these experiments was that text summaries consistently improved retention for summarized content. In addition, summaries presented before and after a text were found to be equally effective.

In view of these considerations, we conducted two follow-up studies (H. van der Meij & van der Meij, 2016a, 2016b) to investigate the effectiveness of reviews. The content and design of the videos in these studies were again comparable to what was used in the earlier research. These studies included an experimental (demonstrations with reviews) and control condition (demonstrations only). In both studies, comparison of task performance showed a statistically significant advantage for the review condition. In the first study, mean success rates for aided task performance were high for both the experimental and control conditions (88% and 86%, respectively). Participants in the review condition outperformed those in the control condition on unaided task performance, with a mean success rate (86%) that approximated the training result, whereas those in the control condition showed a drop (77%). In the second study, participants in the review condition outperformed control condition participants on both aided and unaided task performance. This experiment differed from the first in the absence of user control. Video play was strictly controlled by the experimenter, with each demonstration (and review) video being shown only once to the whole group via an Interactive Whiteboard. Not surprisingly, this restricted viewing led to substantially lower absolute scores on task performances.

Both experiments also found that the training significantly increased the user's motivation (i.e., task relevance and self-efficacy). In addition, the first study found that the review condition increased task relevance perceptions more than the control, while the second study found that the review condition yielded a higher increase in self-efficacy than the control.

4. Experimental design and research questions

The current empirical study further investigated the effectiveness of reviews in instructional videos for software training. The study consisted of two conditions. The experimental condition included a tutorial with demonstration and review videos for each task. In contrast, the control condition presented a tutorial with demonstration videos only. Four research questions were investigated.

Research question 1: How engaging are the videos? Earlier research did not record engagement. This is an omission because it is a prerequisite for learning. This study looked at coverage and total play time as a measure of engagement. Coverage refers to the percentage of a video that was set in play mode. Total play time is the total amount of time that a video was set in play mode. For both measures, no differences between conditions were expected.

Research question 2: How well do the videos support self-efficacy development? Earlier research showed that the videos increased the users' self-efficacy beliefs significantly and substantially. A similar effect was expected in the present study. It was hypothesized that the gain in self-efficacy would be higher in the review condition.

Research question 3: How effective are the videos? Earlier research showed that the videos significantly increased the users' rate of successful task performance. A similar effect was expected in the present study. In addition, it was hypothesized that there would be a stronger increase in the review condition because of the contribution reviews make to the retention process.

Research question 4: What is the relationship between the final appraisal of self-efficacy and success on the post-test? This question investigated whether students' final self-efficacy was aligned with their post-test score. Based on the literature (Bandura, 2012; Komarraju & Nadler, 2013; Stajkovic & Luthans, 1998), it was hypothesized that there would be a significant, positive relationship.

5. Method

5.1. Participants

Seventy-seven participants took part in the study. Three participants from this group were excluded because they did not take the pre-test; one participant was excluded because of missing engagement data. The participants came from three classrooms. One classroom was students in the sixth (and highest) grade of an elementary school. The two other classrooms were students in the first grade in secondary education. The mean age of the 36 male and 37 female participants was 12.8 years. Participants were randomly assigned to the control or experimental condition, after stratification for school and gender. All instructional materials, including the software, were in the participants' native language.

5.2. Instructional materials

Video tutorial. The tutorial used revolved around formatting texts in Microsoft Word. The tasks that were taught were anchored in the user's task domain (M1). The principle of domain anchoring has been advocated by the minimalist approach to software training (H. van der Meij & Carroll, 1998). More generally, one of Merrill's (2002) first principles of design is for instructions to revolve around tasks that are interesting, relevant and engaging for the audience. In the videos, domain anchoring was achieved by instructing participants about important features of reports that must regularly be produced for school.

The basic tutorial (for both the experimental and the control conditions) was organized in 'chapters' with paragraphs that gave access to the videos. Covering all of the chapter topics in a single demonstration video would make the instructions too complex. Therefore, each chapter was split into meaningful and manageable video clips. Chapter 1 included two videos, on adjusting the left-hand and right-hand margins of a text document. Chapter 2 included four videos, on formatting paragraphs, citations, and lists. Chapter 3 presented three videos, on creating an automatic table of contents.

The segmentation of the chapters led to relatively short video clips (under 2 min). Empirical research on video length (M4) indicates that shorter videos are more likely to prevent early dropout by viewers (Guo, Kim, & Rubin, 2014; Wistia, 2012). In addition, empirical research has repeatedly found that segmentation (R1) enhances learning from multimedia (e.g., Margulieux, Guzdial, & Catrambone, 2012; Mayer & Pilegard, 2014; Schittek Janda et al., 2005). An instructional feature that is often mentioned together with segmentation is the inclusion of pauses (R4). Two-second pauses were included in the video clips when there was an event boundary. During such pauses no new (visual or auditory) information was provided to the user; the video simply progressed without stating or showing anything novel. Pauses give the user a brief moment to digest the dynamic (transient) nature of video. Multimedia studies have shown that pauses can enhance learning (e.g., Hassanabadi, Robatjazi, & Savoji, 2011; Lusk et al., 2009; Spanjers, van Gog, & van Merriënboer, 2012).

The presentation of the chapters and clips followed a simple-to-complex sequence (R3) (van Merriënboer & Sweller, 2005). For instance, the clip on adjusting the left margin appeared before the clip on adjusting the right margin, because it involved a similar but slightly easier positioning of the cursor.

The tutorial was presented on a website (see Fig. 2). The left-hand section containing the menu with the table of contents was permanently visible. Clicking on a chapter title led to presentation of a list of the corresponding formatting tasks addressed within that chapter. Clicking on a paragraph title changed its color, and made the task demonstration or review video available. The top of the video screen section on the right displayed the paragraph title, and the type of video (i.e., demonstration or review). To start the video, the participant pressed a prominent play button. When the video started playing, a toolbar appeared that afforded user control (A4) with the usual options of play, pause, resume and stop. Empirical studies have generally reported positive effects of user control on learning from dynamic visualizations (e.g., Berney & Bétrancourt, 2016; Höffler & Schwartz, 2011; Merkt & Schwan, 2014).

Construction of the core component of task demonstration was based on two important design considerations concerning the choice of solution method and nature of the visual and auditory presentation. Software programs usually provide multiple options for how to complete a task. The designer of an instructional video must therefore choose what method(s) to instruct. For a tutorial, the advice is to present only a single method and to choose the one that is easiest to accomplish (Price & Korman, 1993; Renkl, 2014; Smith & Ragan, 2005). For this reason, the tutorial used in the present study consistently presented a single, menu-based method. Furthermore, as advised, it included animated screen recordings with narration (Plaisant & Shneiderman, 2005). Multimedia research has cautioned against using a computer-generated voice for the narration (Mayer, 2014b). The narration in the videos therefore came from a real person (i.e., a native, female speaker).

The *demonstration videos* modeled task completion in a step-by-step fashion. The first part of the narrative supported goal setting. A Word file was shown on the screen and the narrator drew the participants'attention to the formatting problem that it presented. Thereafter, the demonstrations displayed the correct task procedure to participants. Signaling (A1) techniques such as zooming-in and highlighting drew attention to vital screen objects or areas. Eye-movement studies have shown that signals tend to have an attention-drawing effect in multimedia design (e.g., Boucheix & Lowe, 2010; Ozcelik, Arslan-Ari, &

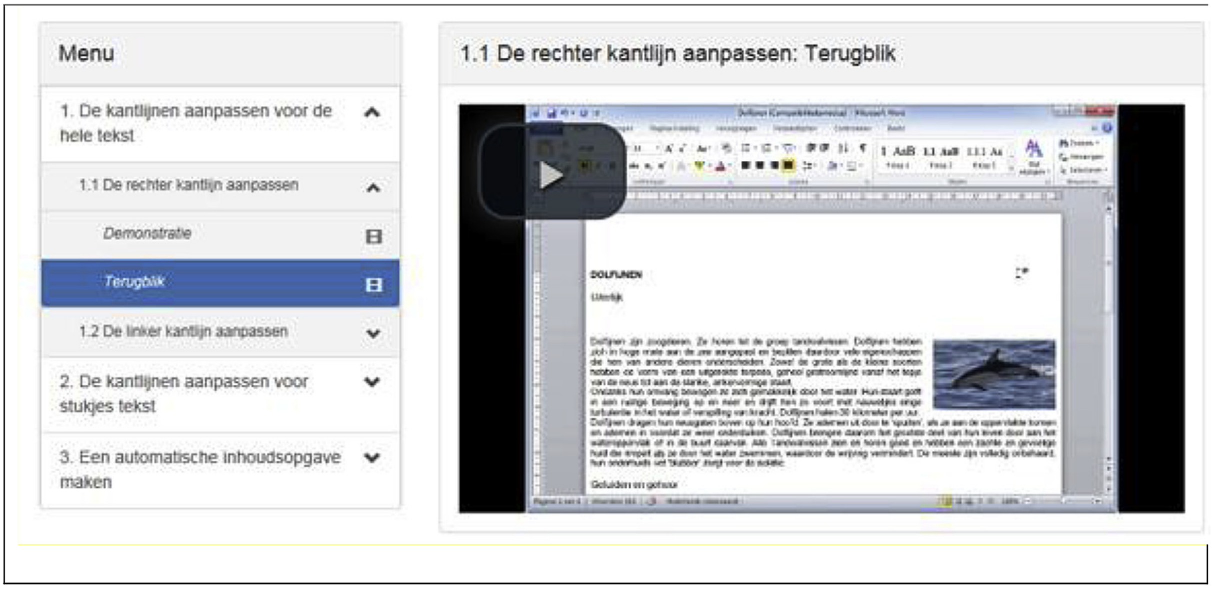

Fig. 2. Screenshot of the website for the video tutorial.

Cagiltay, 2010; de Koning, Tabbers, Rikers, & Paas, 2010) and that their presence generally enhances learning (e.g., Amadieu, Mariné, & Laimay, 2011; Jin, 2013; Richter, Scheiter, & Eitel, 2015).

Action steps were narrated in the form of commands (e.g., “Click the left mouse button”, or “Drag the margin to 2.5 cm”). This is a recommended form of address for action steps in software instructions (Farkas, 1999; Price & Korman, 1993). In contrast, there were personalized messages when the narrative drew participants' attention to screen changes (e.g., “You now see a dotted line”, and “You can see that the margin has become wider”). Multimedia research has frequently indicated that usage of personal pronouns helps create a conversational (versus formal) style (M2) that contributes to motivation as well as learning (e.g., Mayer, Fennell, Farmer, & Campbell, 2004; Reichelt, Kämmerer, Niegemann, & Zander, 2014). Recent empirical research has suggested that culture and the duration of the training can reduce the effectiveness of this design feature (e.g., Brom, Bromová, Děchtěrenko, Michaela Buchtová, & Pergel, 2014; Ginns, Martin, & Marsh, 2013). The pace (A3) of the videos was largely dictated by the narrative. The female voice-over went at a normal speaking rate, yielding a mean words-per-minute (wpm) count of 125. This falls within the range of 125–150 wpm found to be the average rate of human speech (Fulford, 1992). The mean length of the task demonstration videos was 1.14 min (range 0.43–1.46). The total length of all demonstrations was 8.15 min.

The *review videos* (R5) (re)modeled task completion in a more succinct form. The task instructions in the narrative were framed to align with the user's presumed mental rehearsal. That is, they were personalized to take an “I” perspective (e.g., “I must click on …”, and “I should select …”). The mean length of the review videos was 25 s (range 20–36). The total length of all reviews was 3.48 min.

Practice files. Production processes could not be directly supported by the videos. Therefore, the experimental procedure incorporated periods of practice (P1) during training. There was a built-in moment to practice a task immediately after the video(s) on that task. This practice timing (P2) optimized the chances of successful task achievement during training, but could be less effective for supporting learning (Helsdingen, van Gog, & van Merriënboer, 2011). During practice, the participants worked on practice files (P3) that were visibly (but not structurally) different from the example document in the video. Practice files contribute to the effectiveness and efficiency of the user's hands-on practice (H. van der Meij & Carroll, 1998). Tasks completed when working with the practice files were scored in the same way as in the performance tests (see section 5.3 below) to obtain a performance measure during training.

Instruction booklet. A paper booklet guided participants through the experimental procedure by telling them when to play a video, and when to switch to practice. The participants in the control condition were always directed to play the task demonstration first and then engage in practice on that task. The participants in the experimental condition were directed to play the task demonstration first and then play the review, before being invited to practice. The booklet followed a task sequence that was identical with the website's table of contents.

5.3. Measurement instruments

Performance tests. There were three nearly identical performance tests (i.e., pre-test, training, and post-test). Each test included items for all six main formatting tasks in the videos. During testing, participants manipulated dedicated Word files. These files all differed in surface features, but had the same underlying structure (i.e., formatting problem). A score of 0 points was awarded for a task that was not completed correctly. Correct task completion yielded a score of 1. Two formatting tasks (i.e., citations and lists) were judged on the achievement of two distinct sub-goals, yielding a maximum score of 8 for each test. Scores were converted to percentage of possible points. Reliability analyses using Cronbach's alpha led to satisfactory results for the three performance tests: pre-test ($\alpha = 0.67$), training ($\alpha = 0.66$), and post-test ($\alpha = 0.68$).

Self-efficacy Questionnaires. Self-efficacy was assessed with a paper-and-pencil questionnaire administered before and after training. These questionnaires included items addressing each of the six main formatting tasks. Participants responded to statements about formatting (e.g., “I can present a nicely structured list”, and “I know how to indent the first line of a new text segment”) on a 7-point Likert scale. Following Hartley and Betts (2013), positive responses were placed to the left. That is, the scale values ranged from very well (1) to very poorly (7). These scores were reversed for data presentation (i.e., a higher number refers to higher self-efficacy) because such values are easier to read. Reliability analyses using Cronbach's alpha led to satisfactory results for initial ($\alpha = 0.71$), and final self-efficacy ($\alpha = 0.77$).

User logs. A logging program connected to a database was employed to gather activity statistics (i.e., playing, stopping, pausing, or rewinding) for each video second. A popular engagement measure in data mining is “total time engagement”. This measure typically includes moments of video play, but also pauses, replays and rewinds (see Guo et al., 2014; Wistia, 2012). The present study assessed engagement in two slightly different ways.

First, because we were interested in how much of the video was played and potentially viewed, we measured coverage, which we took as the number of video seconds that were presented at least once in play mode. It was expressed as a percentage of the total number of seconds in the video. The length of each video formed the baseline. To illustrate, the length of the video for adjusting the right-hand margin was 55 seconds (i.e., s). When the user log revealed that the number of unique (non-overlapping) play moments was 40 s, coverage was computed to be 73% (40/55). Coverage thus signaled whether a video was played, and therefore potentially viewed, in its entirety. A 0% coverage score indicated that the video was never set in play mode. A score of 100% was the maximum, indicating that each and every video second was played at least once.

The second measure of engagement was total play time. This measure takes the total number of seconds of video played as a percentage of the total number of seconds in all of the videos (477 s for demonstrations, and 255 s for reviews). Any score above 100% indicates that some sections of the videos were played more than once. For instance, if the user log revealed that demonstration videos were played for a total of 527 s, total play time would be 110% (527/477).

5.4. Procedure

The experiment was conducted in two sessions that took place in the schools' computer rooms. The first session began with a 5-minute introduction of the study, after which the initial self-efficacy questionnaire and pre-test were completed (20 min). Training took place one or two days later with separate sessions for the control and experimental groups. Each training began with a whole group introduction (10 min) in which participants were made acquainted with the website, video types and access. Usage of the instruction booklet was also explained. Participants were further instructed to play or replay sections of a video until they felt sure they could complete the task. They were told that they were not allowed to return to a task video once they started the practice item for that task. (User logs revealed that such look-backs were scarce.) Participants were told to work independently and to call for assistance only when stuck. During training, the participants wore headphones. The maximum time for training was 45 min. After a 5-minute break, the final self-efficacy questionnaire was adminstered (3 min). It was followed by the post-test (20 min).

5.5. Analysis

A check on the random distribution of participant characteristics such as age, gender, motivation and prior knowledge (i.e., pre-test) revealed no statistically significant differences for conditions. Tests on assumptions on normality of distribution, and homogeneity of variance revealed violations for dependent variables (i.e., coverage, self-efficacy, performance tests). Therefore, we reported the outcomes from non-parametric tests (i.e., Mann-Whitney *U* test, and Wilcoxon matched-pairs signed rank T-test). The degrees of freedom reported for these findings occasionally differ due to missing data. Due to miscommunication, the initial motivation questionnaire was administered to only a subsample of participants. Tests in the predicted direction were one-tailed, with alpha set at 0.05.

6. Results

6.1. Engagement

Table 1 shows the findings for coverage, organized by condition and video type. In both conditions, the mean coverage for demonstrations was above ninety percent, virtually the same across conditions. Furthermore, analysis of the findings for the separate task demonstrations revealed that mean coverage scores ranged from a low of 85.1% to a high of 98%. In short, the findings show that most participants chose to play most of the demonstrations in the tutorial.

Mean coverage for the review videos was just over 30%. The difference in coverage between reviews and demonstrations was statistically significant, $T(36) = 5$, $p < 0.001$. Analysis of separate review videos further revealed considerable differences for mean coverage, ranging from a low of 17.4% to a high of 67.4%. The highest coverage was for the first task in the tutorial; all later review videos had mean scores of 37% or below.

The total play time for all demonstrations in both conditions was just over 100%. For the demonstrations, the control condition had a total play time of 102% ($SD = 15.0$), while the experimental condition had a score of 100% ($SD = 22.2$). The difference between conditions was not statistically significant. The total play time for all reviews was 30% ($SD = 35.2$). The difference in total play time between reviews and demonstrations was statistically significant, $F(1,35) = 130$, $p < 0.001$.

6.2. Self-efficacy

Data about initial self-efficacy, available from only one classroom, indicated that participants began training with a mean score of 3.61 ($SD = 1.45$), which is slightly below the scale mean value of 4. After training, the mean rose to 6.45 ($SD = 0.71$). This was a statistically significant improvement, $T(21) = 231$, $p < 0.001$. Participants in the review condition ended training with higher self-efficacy scores ($n = 36$, $M = 6.63$, $SD = 0.45$) than those in the control condition ($n = 34$, $M = 6.27$, $SD = 0.88$). This difference was not statistically significant.

Table 1
Mean coverage (standard deviation) per condition and video type.

Condition	Demonstrations		Reviews	
	Mean	(SD)	Mean	(SD)
Control ($n = 37$)	94.1%	(10.7)	n.a.	
Review ($n = 36$)	92.0%	(13.9)	32.5%	(34.7)

6.3. Performance test scores

Table 2 shows that the success rate on the pre-test was just under 30%. During training this rose to an average success rate of 84%, which is a statistically significant gain over the pre-test score, $T(72) = 2.481$, $p < 0.001$. Conditions differed for task performance scores during training. Comparison of these success rates showed a statistically significant advantage for participants in the review condition over those in the control condition, $U(72) = 801$, $p = 0.035$ (one-sided).

The average success rate on the post-test was 71%. Compared to the results obtained during training, this was a statistically significant decrease, $T(65) = 190{,}5$, $p < 0.001$. However, compared with the pre-test scores, the difference was still positive and statistically significant, $T(67) = 2.056$, $p < 0.001$. Conditions also differed for scores on the post-test. Participants in the review condition had a statistically significant higher success rate on this test than did those in the control condition, $U(67) = 699$, $p = 0.038$ (one-sided).

6.4. Self-efficacy and post-test

There was a positive and statistically significant relationship between final self-efficacy and post-test score, $r(66) = 0.395$, $p = 0.001$. This finding shows that participants who expressed higher confidence in successful task completion of trained tasks after training also did better on the subsequent final performance test than participants with lower post-training self-appraisals.

7. Conclusion and discussion

The coverage data clearly show that the demonstration videos engaged the participants. All of these videos were played almost completely, regardless of whether they appeared at the start, in the middle or at the end of the tutorial. The experimental setting of the study may have contributed to the high level of coverage, but presumably the design of the tutorial is also to be credited. The theoretical model suggests that the instructional features related to affective outcomes are primarily responsible for high engagement with the demonstrations.

The coverage data for reviews show that engagement with this type of video was meager to poor. While the first review still received fairly high coverage, immediately thereafter interest waned. It seems reasonable to assume that this signals that participants actively engaged with the first task review to get acquainted with what new information this type of video had to offer. The coverage data suggest that the first impression was enough to make the participants decide that reviews did not need to be played in full, or could be ignored altogether.

The meager to poor coverage for reviews may have at least three different grounds. One, the demonstration videos were designed with instructional support that addressed key processes and conditions in observational learning. The demonstrations may thus have given the participants a sufficiently clear view of what they needed to learn, obviating the need for reviews. Two, task execution in the review used the same text file as in the demonstration. This has the advantage that it gives the participant the impression that nothing new was being taught, that the review involved the same task completion in condensed form. Perhaps using a different file for the formatting task in the review would have led the participants to keep on playing it longer. The variation might raise motivation, but it might decrease transparency of the link between demonstration and review. Three, the reviews were about one-third the length of the demonstrations. This appears to be relatively long, especially when compared to the upper limit of 10% advocated for text summaries (e.g., Adelaide, 2014; Driscoll, 2013). In other words, the reviews may have been a bit too long, causing early drop-out.

The mean total play time for demonstrations was about 100% and virtually the same in both conditions. In other words, these data reveal that the average participant rarely replayed a section of a demonstration. Demonstrations tended to be played once and almost in full, regardless of condition. In other words, participants in the control condition were not more likely to 'compensate' for the absence of reviews by replaying sections of the demonstrations more often than participants that had access to the reviews.

A significant increase in self-efficacy was found. Participants started with a moderate level of confidence in their capacity to complete the to-be-trained tasks. After training, they had become much more confident. The performance data generally corroborate this change in appraisal. Further support comes from the finding of a positive and significant correlation between self-efficacy and post-test score. The gain in self-efficacy is important in itself, but it is also relevant for future actions, as high

Table 2
Mean performance (standard deviation) per condition and test type.

Condition	Pre-test		Training		Post-test	
	Mean	(SD)	Mean	(SD)	Mean	(SD)
Control ($n = 37, 36, 35$)[a]	29.4%	(24.9)	80.9%	(20.0)	65.4	(27.6)
Review ($n = 36, 36, 32$)[a]	29.2%	(21.5)	87.9%	(18.6)	77.1	(19.2)
Total ($n = 73, 72, 67$)[a]	29.3%	(23.1)	84.4%	(19.5)	71.3	(24.6)

[a] The number of participants for pre-test, training and post-test.

self-efficacy has been found to relate to people trying harder and being more persistent with comparable task challenges (e.g., Bandura & Locke, 2003; Bandura, 2012).

The prediction that final self-efficacy would be higher in the experimental condition was not confirmed. One reason for this finding might be that there was no restriction on replaying the demonstration video or time spent on practice. Participants in the control group could view the demonstration videos or practice the tasks until they felt self-efficacious. This might explain why self-efficacy was same for both experimental conditions.

The outcomes for the performance tests indicate that the videos effectively supported task completion and learning. Participants started training with a success rate below 30%. During training this increased to over 80%. Task completion when practicing during training was unaided; participants were expected not to consult the task video anymore. In short, the findings for practice tasks indicate learning. During training, the instructions for a specific formatting task immediately preceded practice on that task. Participants probably benefitted from this proximity. In the post-test there was no such alignment. Participants then had to face the additional challenge of selecting the proper method for each formatting task. In other words, in addition to testing for recall of the steps in a procedure, the post-test called for knowledge of selection rules (Card, Moran, & Newell, 1983). The data show that success rates on the post-test were indeed lower than during training, but the average of 71% was still substantially higher than the pre-test score.

The review condition did significantly better than the control on the performance tests during and after training. These outcomes confirm our prediction, but given the meager to poor coverage scores for reviews, this result is surprising. Does it signal, for instance, that maybe the review served as a frame of reference for comparison to the user's own summary? Perhaps simply beginning to watch the review prompted the users' mental rehearsal of the necessary steps, and when no discrepancy came up immediately, the user felt confident of having a correct understanding stored in memory (and therefore did not feel the need to watch the entire review). And how does this outcome relate to the reviews' presumed compensation for mind wandering during the task demonstrations? And what about the possibility that reviews served as a repetition? The present study gives no answer to these questions about the contribution of reviews to learning. It also calls for a more precise account of what happens in retention when a user tries to commit task performance to long-term memory.

The present study represents only a first step in building understanding of what it takes to learn from observing a model of performance and how to support the processes involved. We have found the theoretical model quite useful for constructing instructional videos for software training. An important advantage of this framework over other research on design principles for video construction (e.g., Koumi, 2013; H.; van der Meij & van der Meij, 2013) is the linkage with key observational learning processes.

The results are promising. Motivation improved and task performance increased substantially after the demonstrations, and task performance increased even more so after a combination of demonstration and review videos. The findings are also in line with the outcomes of comparable earlier studies (H. van der Meij & van der Meij, 2016a, 2016b). Future research might want to delve deeper into understanding how instructional features affect observational processes and what users remember from an instructional video.

Acknowledgement

The author wishes to thank Bas Kolloffel and Elke Hartman for their help in conducting the experiment.

References

Adelaide, T. U. o (2014). *Writing an abstract*. Writing Centre Learning Guide. Retrieved from http://www.adelaide.edu.au/writingcentre/learning_guides/learningGuide_writingAnAbstract.pdf.

Adobe. (2017). *Premiere Pro tutorials*. Retrieved from https://helpx.adobe.com/premiere-pro/tutorials.html.

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. http://dx.doi.org/10.1016/j.chb.2010.05.009.

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

Bandura, A. (2012). On the functional properties of perceived self-efficacy revisited. *Journal of Management, 38*(1), 9–44. http://dx.doi.org/10.1177/0149206311410606.

Bandura, A., & Locke, E. A. (2003). Negative self-efficacy and goal effects revisited. *Journal of Applied Psychology, 88*(1), 87–99. http://dx.doi.org/10.1037/0021-9010.88.1.87.

Berney, S., & Bétrancourt, M. (2016). Does animation enhance learning? A meta-analysis. *Computers & Education, 10*, 150–167. http://dx.doi.org/10.1016/j.compedu.2016.06.005.

Boucheix, J. M., & Lowe, R. (2010). An eye tracking comparison of external pointing cues and internal continuous cues in learning with complex animations. *Learning and Instruction, 20*, 123–135. http://dx.doi.org/10.1016/j.learninstruc.2009.02.015.

Brar, J., & van der Meij, H. (2017). Complex software training: Harnessing and optimizing video instructions. *Computers in Human Behavior, 70*, 1–11. http://dx.doi.org/10.1016/j.chb.2017.01.014.

Brom, C., Bromová, E., Děchtěrenko, F., Michaela Buchtová, M., & Pergel, M. (2014). Personalized messages in a brewery educational simulation: Is the personalization principle less robust than previously thought? *Computers & Education, 72*, 339–366. http://dx.doi.org/10.1016/j.compedu.2013.11.013.

Card, S. K., Moran, T. P., & Newell, A. (1983). *The psychology of human-computer interaction*. Hillsdale, NJ: Erlbaum.

Carroll, J. M., & Rosson, M. B. (1987). The paradox of the active user. In J. M. Carroll (Ed.), *Interfacing thought: Cognitive aspects of human-computer interaction* (pp. 80–111). Cambridge, MA: MIT Press.

Christensen, C. M., & Stordahl, K. E. (1955). The effect of organizational aids in comprehension and retention. *Journal of Educational Psychology, 46*(1), 65–74.

Driscoll, D. L. (2013). *Writing report abstracts*. Owl Purdue Online Writing Lab. Retrieved from https://owl.english.purdue.edu/owl/resource/656/1/.

Farkas, D. K. (1999). The logical and rhetorical construction of procedural discourse. *Technical Communication, 46*, 42–54.

Fu, W. T., & Gray, W. D. (2004). Resolving the paradox of the active user: Stable suboptional performance in interactive tasks. *Cognitive Science, 28*(6), 901–935. http://dx.doi.org/10.1016/j.cogsci.2004.03.005.

Fulford, C. P. (1992). *Systematically designed text enhanced with compressed speech audio. Paper presented at the Annual Meeting of the Association for Educational Communications and Technology*. Washington: D.C.

Gil, L., Bråten, I., Vidal-Abarca, E., & Strømsø, H. I. (2010). Summary versus argument tasks when working with multiple documents: Which is better for whom? *Contemporary Educational Psychology, 35*, 157–173. http://dx.doi.org/10.1016/j.cedpsych.2009.11.002.

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. http://dx.doi.org/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 & Education, 12*(2), 219–243. http://dx.doi.org/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. In *Paper presented at the L@S '14*. Atlanta, GA.

Hartley, J., & Betts, L. (2013). Let's be positive: The effects of the position of positive and negative values and labels on responses to Likert-type scales. *Chinese Journal of Psychology, 55*(2), 291–299. http://dx.doi.org/10.6129/CJP.20130319.

Hartley, J., Goldie, M., & Steen, L. (1976). The role and position of summaries: Some issues and data. *Educational Review, 31*(1), 59–65. http://dx.doi.org/10.1080/0013191790310107.

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

Hassanabadi, H., Robatjazi, E. S., & Savoji, A. P. (2011). Cognitive consequences of segmentation and modality methods in learning from instructional animations. *Procedia - Social and Behavioral Sciences, 30*, 1481–1487. http://dx.doi.org/10.1016/j.sbspro.2011.10.287.

Helsdingen, A., van Gog, T., & van Merriënboer, J. J. G. (2011). The effects of practice schedule and critical thinking prompts on learning and transfer of a complex judgment task. *Journal of Educational Psychology, 103*(2), 383–398. http://dx.doi.org/10.1037/a0022370.

Höffler, T. N., & Schwartz, R. N. (2011). Effects of pacing and cognitive style across dynamic and non-dynamic representations. *Computers & Education, 57*, 1716–1726. http://dx.doi.org/10.1016/j.compedu.2011.03.012.

Ikea. (2017). *Ikea Pax Wardrobe with hinged doors assembly instructions*. Retrieved from https://www.youtube.com/watch?v=tTeI1FSBqYw.

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

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review, 19*, 509–539. http://dx.doi.org/10.1007/s10648-007-9054-3.

Kim, J., Nguyen, P., Weir, S., Guo, H. J., Miller, R. C., & Gajos, K. Z. (2014). Crowdsourcing step-by-step information extraction to enhance existing how-to videos. In *Paper presented at the CHI 2014* (Toronto, ON, Canada).

Komarraju, M., & Nadler, D. (2013). Self-efficacy and academic achievement: Why do implicit beliefs, goals, and effort regulation matter? *Learning and Individual Differences, 25*(1), 67–72. http://dx.doi.org/10.1016/j.lindif.2013.01.005.

de Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2010). Learning by generating vs. receiving instructional explanations: Two approaches to enhance attention cueing in animations. *Computers & Education, 55*, 681–691. http://dx.doi.org/10.1016/j.compedu.2010.02.027.

Kosoff, M. (2017). *LinkedIn just bought online learning company Lynda for $1.5 billion*. Retrieved from http://www.businessinsider.com/linkedin-buys-lyndacom-for-15-billion-2015-4?international=true&r=US&IR=T.

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. http://dx.doi.org/10.1080/23796529.2013.11674711.

Leopold, C., Sumfleth, E., & Leutner, D. (2013). Learning with summaries: Effects of representation mode and type of learning activity on comprehension and transfer. *Learning and Instruction, 27*, 40–49. http://dx.doi.org/10.1016/j.learninstruc.2013.02.003.

Lusk, D. L., Evans, A. D., Jeffrey, T. R., Palmer, K. R., Wikstrom, C. S., & Doolittle, P. E. (2009). Multimedia learning and individual differences: Mediating the effects of working memory capacity with segmentation. *British Journal of Educational Technology, 40*(4), 636–651. http://dx.doi.org/10.1111/j.1467-8535.2008.00848.x.

Margulieux, L., Guzdial, M., & Catrambone, R. (2012). Subgoal-labeled instructional material improves performance and transfer in learning to develop mobile applications. In *Paper presented at the 9th international conference on international computing education research (ICER'12)* (Auckland, New Zealand).

Mayer, R. E. (2014a). *The Cambridge handbook of multimedia learning* (2nd ed.). New York, NY: Cambridge University Press.

Mayer, R. E. (2014b). Principles based on social cues in multimedia learning: Personalization, voice, image and embodiment principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 345–368). New York, NY: Cambridge University Press.

Mayer, R. E., Fennell, S., Farmer, L., & Campbell, J. (2004). A personalization effect in multimedia learning: Students learn better when words are in conversational style rather than formal style. *Journal of Educational Psychology, 95*, 389–395. http://dx.doi.org/10.1037/0022-0663.96.2.389.

Mayer, R. E., & Pilegard, C. (2014). Principles for managing essential processing in multimedia learning: Segmenting, pre-training, and modality principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 316–344). New York, NY: Cambridge University Press.

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

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

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* (pp. 19–53). 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*(3), 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. http://dx.doi.org/10.1016/j.compedu.2014.06.003.

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*(2), 116–132. http://dx.doi.org/10.1111/jcal.12082.

van der Meij, H., & van der Meij, J. (2016a). Demonstration-Based Training (DBT) for the design of a video tutorial for software instructions. *Instructional Science, 44*, 527–542. http://dx.doi.org/10.1007/s11251-016-9394-9.

van der Meij, H., & van der Meij, J. (2016b). The effects of reviews in video tutorials. *Journal of Computer Assisted Learning, 32*, 332–344. http://dx.doi.org/10.1111/jcal.12136.

Merkt, M., & Schwan, S. (2014). How does interactivity in videos affect task performance? *Computers in Human Behavior, 31*, 172–181. http://dx.doi.org/10.1016/j.chb.2013.10.018.

van Merriënboer, J. J. G., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review, 17*(2), 147–177. http://dx.doi.org/10.1007/s10648-005-3951-0.

Merrill, M. D. (2002). First principles of instruction. *Educational Technology Research & Development, 50*(3), 43–59. http://dx.doi.org/10.1007/BF02505024.

Microsoft. (2017). *Word training*. Retrieved from https://support.office.com/en-us/article/Word-training-7bcd85e6-2c3d-4c3c-a2a5-5ed8847eae73?ui=en-US&rs=en-US&ad=US.

Ozcelik, E., Arslan-Ari, I., & Cagiltay, K. (2010). Why does signaling enhance multimedia learning? Evidence from eye movements. *Computers in Human Behavior, 26*(1), 110–117. http://dx.doi.org/10.1016/j.chb.2009.09.001.

Plaisant, C., & Shneiderman, B. (2005). Show me! Guidelines for recorded demonstration. In *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. http://dx.doi.org/10.1016/j.chb.2011.12.001.

Price, J., & Korman, H. (1993). *How to communicate technical information*. Redwood City, CA: The Benjamin/Cummings Publishing Company, Inc.

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. http://dx.doi.org/10.1016/j.chb.2014.03.005.

Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cognitive Science, 38*, 1–37. http://dx.doi.org/10.1111/cogs.12086.

Richter, J., Scheiter, K., & Eitel, A. (2015). Signaling text-picture relations in multimedia learning: A comprehensive meta-analysis. *Educational Research Review, 17*, 19–36. http://dx.doi.org/10.1016/j.edurev.2015.12.003.

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*(5), 596–609. http://dx.doi.org/10.1177/0018720810381071.

Schittek Janda, M., Botticelli, T. A., Mattheos, N., Nebel, D., Wagner, A., Nattestad, A., et al. (2005). Computer-mediated instructional video: A randomised controlled trial comparing a sequential and a segmented instructional video in surgical hand wash. *European Journal of Dental Education, 9*, 53–58. http://dx.doi.org/10.1111/j.1600-0579.2004.00366.x.

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

Spanjers, I. A. E., van Gog, T., & van Merriënboer, J. J. G. (2012). Segmentation of worked examples: Effects on cognitive load and learning. *Applied Cognitive Psychology, 26*, 352–358. http://dx.doi.org/10.1002/acp.1832.

Stajkovic, A. D., & Luthans, F. (1998). Self-efficacy and work-related performance: A meta-analysis. *Psychological Bulletin, 124*(2), 240–261. http://dx.doi.org/10.1037/0033-2909.124.2.240.

TechSmith. (2017). *Camtasia 9.3 tutorials*. Retrieved from https://www.techsmith.com/tutorial-camtasia-9-3.html.

Vezin, J. F., Berge, O., & Mavrelis, P. (1973). Role du résumé et de la répétition en fonction de leur place par rapport en texte. *Bulletin de Psychologie, 27*, 163–167.

Westby, C., Culatta, B., Lawrence, B., & Hall-Keyon, K. (2010). Summarizing expository texts. *Topics in Language Disorders, 30*(4), 275–287. http://dx.doi.org/10.1097/TLD.0b013e3181ff5a88.

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