AECT | ASSOCIATION FOR EDUCATIONAL COMMUNICATIONS & TECHNOLOGY

CrossMark

RESEARCH ARTICLE

Supporting motivation, task performance and retention in video tutorials for software training

Hans van der Meij[1] [ID] · **Jan van der Meij[2]** · **Tessa Voerman[1]** ·
Evert Duipmans[3]

Abstract Video tutorials for software training are becoming more and more popular, but their construction and effectiveness is understudied. This paper presents a theoretical model that combines demonstration-based training (DBT) and multimedia learning theory as a framework for design. The study investigated the effects of video tutorials on motivation, task accomplishment and learning with special attention being paid to the role of reviews. Three tutorials were compared: preview and demonstration (control condition), preview and demonstration and review (review condition) and preview and demonstration and 2nd demonstration (2-demo condition). Participants were 65 students from the upper grades of elementary school. Participant logs revealed a significant decrease in coverage for video types following the preview. All tutorials significantly raised self-efficacy and procedural knowledge (i.e., task performances during training, immediate and delayed post-test), but no effects of condition were found. The discussion addresses issues concerning the design of the tutorials (overlap between video types) and usage conditions (blocked video access during training tests) that may have negatively affected coverage and the effectiveness of the reviews. A limitation of the study is that no process data were

✉ Hans van der Meij
  h.vandermeij@utwente.nl

  Jan van der Meij
  j.vandermeij@utwente.nl

  Tessa Voerman
  tevoerman@gmail.com

  Evert Duipmans
  evert@duipmans.nl

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

[2]   Department of ELAN Teacher Development, Faculty of Behavioural, Management and Social Sciences, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands

[3]   Saxion, Maarten Harpertsz, Tromplaan 28, 7513 AB Enschede, The Netherlands

gathered that could shed light on the relationships between design features and observational learning processes.

Introduction

The video tutorial is rapidly gaining ground as an instructional means for software training. Companies such as Adobe, IBM, Microsoft, SAP and TechSmith are accommodating their clients' need for instructional support by developing a rapidly growing set of video tutorials. For instance, TechSmith provides a wide range of video tutorials that help expand the user's skills in handling their programs. For their screen capture program SnagIt, the user can find videos on how to accomplish basic tasks, including screen capturing, editing a capture, and media sharing (TechSmith 2017, June 4). A further sign of the rising popularity is the large repository of video tutorials for software training on the website Lynda.com which recently was sold to LinkedIn in the amount of $1.5 billion (Kosoff 2017, May 2). Although video tutorials for software training may be increasingly popular, it is unclear how well they work. There is no published information about the design characteristics and effectiveness of the video tutorials produced by the aforementioned companies.

Empirical studies that have compared video-based to paper-based tutorials for software training have shown mixed results. While Mestre (2012) reported an advantage of paper-based instructions, Alexander (2013) found no differences. Three other studies found the video tutorial significantly more successful than the paper-based version (Lloyd and Robertson 2012; van der Meij and van der Meij 2014, 2015). What this research suggests is that effective videos should capitalize on the affordances of video while also attending to, and possibly even incorporating key qualities of print.

van der Meij and van der Meij (2014, 2015) also reported the presence of a considerable discrepancy between aided and unaided task performances for the video tutorial. Their findings showed was that the demonstrations of task completion were sufficiently clear to enable the users to perform the modeled tasks, but failed to achieve the same level of success for learning. This result was seen as a signal that there was a need for additional support for the user's retention. A similar finding was obtained in another study by van der Meij (2014) involving the same video tutorial.

The findings led to an elaboration of design guidelines and new empirical studies. Specifically, a theoretical model was proposed consisting of a combination of demonstration-based training (DBT) (Grossman et al. 2013) and multimedia learning theory (e.g., Mayer 2014). Subsequently, a video tutorial for software training was designed in accordance to the model and three experiments were conducted to test the effectiveness of including complementary review videos (henceforth reviews) for enhancing learning (Brar and van der Meij 2017; van der Meij and van der Meij 2016a, b). By and large, these studies found positive effects of reviews.

The present study continues this line of research on reviews. First, the paper briefly summarizes the theoretical model as a guide for constructing a video tutorial for software training. Next, a discussion follows concerning the retention process and the benefits of complementing demonstrations with reviews. The remainder of the paper discusses an

experiment on the added value of reviews for a video tutorial constructed using the guidelines outlined by the theoretical model.

A theoretical model for software training

The video tutorial for software training that we designed consisted of a demonstration of task performance that users could mimic and learn from, and that should also be of interest to them. Figure 1 depicts the theoretical model that formed the basis for this construction effort. At the heart of the model, there are the four fundamental processes involved in
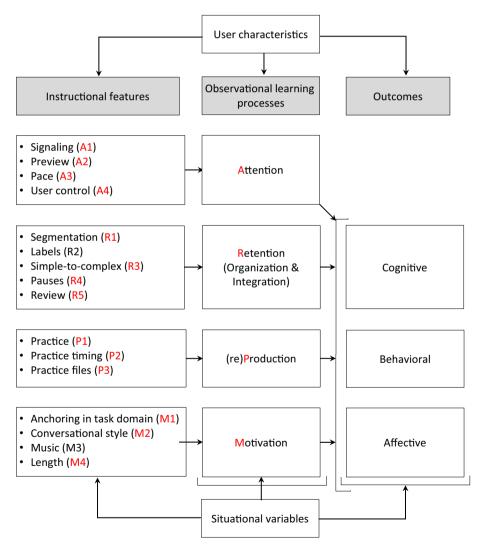

Fig. 1 Theoretical model for video construction for software training. All features, except R2 and M3, were incorporated in the design of the tutorial in this study

observational learning distinguished by Bandura (1986). Connected to these processes are instructional features that, according to demonstration-based training (DBT), are necessary to increase learning (Grossman et al. 2013; Rosen et al. 2010). The instructional features in the model also connect to multimedia learning theory (see Mayer 2014) with its important insights on (limitations of) cognitive processing and large set of empirically tested principles for instructional design.

Under the umbrella term DBT, the model has recently been described in detail (Brar and van der Meij 2017; van der Meij and van der Meij 2016a, b). The next section briefly describes the retention process and summarizes the research on reviews. The subsection on "Instructional materials" in the "Method" details how specific instructional features from the theoretical model found their way into the design of the video tutorial.

Notice that the theoretical model also takes into account user characteristics and situational variables. A user characteristic that should always be considered when seeking to optimize learning is prior knowledge. Research on the expertise reversal principle has shown that design features that are effective for novices or beginners may not be effective or may even hinder learning for more knowledgeable people (Kalyuga 2007, 2014). Situational variables also constrain the design guidelines that can be applied. For software training, the prototypical context is the individual user working alone at his or her computer. Accordingly, the model does not mention various types of instructor-led support.

Complementing a demonstration with a review video

Retention involves the process of transforming incoming information into symbolic codes that are stored in long-term memory (Bandura 1986). The user must extract the distinctive features and structures from a task demonstration and code these into a succinct, prototypical representation. This information must be committed to memory to serve as a guide for subsequent action. Retention, thus, subsumes the processes that Mayer (2008) refers to as organization and integration. The end result should be a concept of task performance that can serve as a guide and standard, enabling the user to organize, initiate, and monitor future actions (Bandura 1986).

Figure 1 shows various design features that can support retention. The original video tutorial used in van der Meij and van der Meij's (2014, 2015) studies incorporated several of these features in its design (i.e., segmentation, simple-to-complex sequencing, and pauses). Still, additional support is needed to enhance user's retention. Including a review feature, which summarizes relevant information of a task performance, in a video tutorial might strengthen the user's retention, recall and task performance.

A literature search of empirical studies on how people summarize multimedia presentations showed no hits. Thus, a literature search on text representation and recall has been conducted. According to Meyer (1975, 1984), proficient readers adopt a structure-based strategy to create a complete and coherent representation of an expository text. In such a strategy, the reader identifies the overall top-level structure (e.g., sequence) of the rhetorical relationships between the main ideas in a text. This structure is then used to organize comprehension and recall. Readers who adopt a structure-based strategy have better understanding and recall of the main topics of a text (Meyer and Poon 2001). Moreover, they are in a better position to recall subordinate content as well (Meyer 1975). Less proficient readers do not automatically use a structure-based strategy, but they benefit from scaffolds that reveal the rhetorical structure of the text (Meyer and Wijekumar 2007).

Empirical research on the effectiveness of reviews in tutorials has been limited (van der Meij and van der Meij 2016a). A few studies on the contribution of text summaries to recall reveal that text summaries can enhance retention for summarized content (Hartley et al. 1976; Hartley and Trueman 1982; McLaughlin Cook 1981; Vezin et al. 1973). Summaries presented before and after a text have been found equally effective.

Three initial experiments have been conducted to investigate whether reviews in a video tutorial contribute to motivation and learning particularly how users could benefit from reviews in different ways (Brar and van der Meij 2017; van der Meij and van der Meij 2016a, b). First, a review provides the user with a concise overview of the main issues or events in a procedure, which can enhance retention and increase the user's self-efficacy, defined as the confidence in his or her capacity for task accomplishment. Second, a review can serve as a frame of reference for the user's own summary of a procedure. This feature can also strengthen the user's knowledge of the correct task solution and self-efficacy belief. A third presumed benefit of a review is that it offers the user a second chance to learn, compensating for any possible mind wandering during the presentation of a demonstration.

Brar and van der Meij (2017) investigated a video tutorial that taught university student how to conduct a *t* test using the SPSS software program. In the control condition, only demonstrations could be viewed. In the experimental condition, a review was presented after each demonstration. The study only assessed the contribution of reviews to learning outcomes. The main finding was that review condition did better on both the knowledge and performance test, but the difference between conditions was not statistically significant.

van der Meij and van der Meij (2016a, b) also compared a demonstrations only control condition to an experimental condition with demonstrations plus reviews. This research was more similar to earlier studies both for the audience (i.e., upper elementary and/or lower secondary school) and domain of the video tutorials (i.e., formatting tasks in Microsoft Word). In the experiment where participants had unlimited access to the videos during training, reviews enhanced self-efficacy (van der Meij and van der Meij 2016a). In the study where participants had no control over video play, reviews were found to significantly enhance task performance (van der Meij and van der Meij 2016b). Both studies also found that learning was significantly higher in the review condition.

The present study further tests the effects of reviews in video tutorials on user motivation and learning for formatting tasks in Microsoft Word. It differs from earlier research in several ways. First, an important aspect neglected in the two earlier studies, which is taken into account in the present study, is viewing behavior. When a review is added to a demonstration video, the total length of the task video increases. This raises the question whether users are willing to spend additional time viewing the review. Studies in which data mining has been applied to determine drop-out rates for videos clearly show that video length and viewing behavior are related (Guo et al. 2014; Wistia 2012). Their findings indicate that shorter videos keep more users on-board and are viewed more thoroughly than longer videos. In other words, data on viewing behavior can show whether reviews are viewed thoroughly and whether this influences their effectiveness.

Second, the procedure of this study limited user control during training. Participants could view the video as long as they wanted, and they could pause and replay sections as they deemed necessary. However, access to the videos was blocked once they started to practice. In effect, this procedure yielded an assessment of what was learned immediately after viewing a video on a task (i.e., a training test). Third, previews were added to the tutorial, which was expected to yield a pre-training effect. Because a slightly younger and

presumably less knowledgeable audience was targeted, the previews were intended to prevent overload problems when processing the demonstration and review (see Mayer and Pilegard 2014). Fourth, a third condition was included. A new tutorial was created in which a second, identical demonstration followed after the first demonstration. This 2-demo condition controls for the possibility that the influence of reviews is mainly a repetition effect.

Experimental design and research questions

The experiment was set up as a factorial design with a control condition (preview and demonstration), and two experimental conditions (preview and demonstration and review, and preview and demonstration and 2nd demonstration). The study aims to address the following research questions.

Research question 1: How fully are the videos viewed?

Empirical research indicates that video length can have a significant effect on viewing behavior. Longer videos often lead to a reduction in the percentage of a video that is viewed (Guo et al. 2014; Wistia 2012). The videos in the present study differ in length. Therefore, data are gathered about video coverage which is the percentage of a video that is shown in play mode. These data give information about the possible role of video length.

Research question 2: How effective are the videos in increasing self-efficacy and is there an effect of condition?

The video tutorials were designed to support motivation. For motivation we measure self-efficacy which can be defined as a person's expectancy for success on novel tasks (Bandura 1997). In line with the finding of an earlier study (van der Meij and van der Meij 2016a), the increase in motivation is expected to be strongest for the review condition.

Research question 3: How well do the videos support learning and is there an effect of condition?

Earlier studies with video tutorials for the same software tasks found that demonstrations plus reviews increased learning more than demonstrations only (van der Meij and van der Meij 2016a, b). This effect is also expected in the present study.

Method

Participants

The 65 participants in the study came from two fifth-grade and two sixth-grade classrooms from an elementary school in the Netherlands. The mean age of the participants was 11.5 years with a standard deviation of 0.71. The 32 boys and 33 girls were randomly assigned to condition after stratification for classroom. All participants were native

speakers of Dutch. Therefore, all instructional materials, including the software, were in Dutch.

Instructional materials

The videos in this study embodied a great number of instructional features from the theoretical model. The tutorial taught six formatting tasks in Microsoft Word 2010. These tasks were anchored in the participants' task domain (M1 in Fig. 1) by starting with poorly formatted original texts from school reports. The example documents were then changed into well-formatted texts in the videos.

The videos in the tutorial were organized in a table of contents with (sub)titles. The first chapter of the tutorial discussed adjusting the left and right margins for a whole document. The second chapter revolved around formatting paragraphs, citations, and lists. The last chapter dealt with the automatic creation of a table of contents.

The division of chapters and the videos therein represented a meaningful segmentation (R1) of the various formatting tasks. The sequencing of the videos followed the simple-to-complex principle (R3). For instance, the video on adjusting the left margin of a complete document preceded the video on adjusting the right margin because it involved the easiest manipulation of the screen object in both tasks.

For individual videos an upper limit of 3 min was chosen. Empirical research suggests that this is a suitable length (M4) for maintaining user interest and preventing early drop-out (Guo et al. 2014; Wistia 2012). Table 1 shows the length of the video types (1st and 2nd demonstrations have the same duration). The table gives the basic data for computing the length of the video tutorials. Specifically, the length of the control tutorial is simply the sum of all previews and demonstrations, and the length of the review tutorial is the sum of all previews, demonstrations, and reviews. For the 2-demo tutorial the length is the sum of all previews and the demonstrations taken twice. There was a considerable difference in video length for the tutorials. The control tutorial lasts for 12.3 min, the review for 16.1 min, and the 2-demo tutorial is 20.5 min of video.

Table 1 Titles and duration (in s) of video types

Title	Preview	Demonstration	Review
Adjusting the margins for the whole text			
Adjusting the right margin	30	55	20
Adjusting the left margin	29	42	21
Adjusting the margins for a text segment			
Indenting a citation to the left	34	45	22
Indenting a citation to the right	26	44	22
Indenting the first line of a paragraph	21	48	28
Improving a list	33	66	31
Creating an automatic table of contents			
Styling the main headings	30	59	25
Styling the subheadings	21	75	35
Creating an automatic table of contents	30	43	21

Participants could access the video tutorial via a website. The left-hand column of this website presented the table of contents that was permanently visible on the screen. The subtitles gave access to the videos. Clicking on a subtitle changed its color and opened the video in the right-hand column of the website. The subtitle also appeared above the video. There was no transcript of the narrative. The toolbar facilitated user control (A4). Students could start, pause, stop, fast-forward and rewind the video with a toolbar. In addition, they could adjust the sound level and switch between a full and partial screen representation.

The basic set-up of the videos was a combination of animated screen recordings with a narrative spoken by a female native speaker. Three types of video were developed for each task, namely, a preview, a demonstration, and a review. First and second demonstrations were identical videos. The type of video that the users could expect to see was announced in writing (e.g., "preview") in a slide that was displayed two seconds before its start. All of the videos for a task automatically appeared one after the other.

In this study, pace (A3), the ongoing stream of information in a video, was measured as the mean number of words-per-minute (wpm). The wpm count for all videos together yielded a score of 118 wpm. Compared to a mean rate of 125–150 wpm for human speech (Fulford 1992), this qualifies the pace of the videos as slow.

Previews (A2) prepared the user for the upcoming task by showing the start and end state of a formatting task. In addition, previews explained and illustrated task-specific concepts such as margins or paragraphs that the user may not yet know. Finally, previews directed the user's attention to pertinent screen objects and locations for task completion. At the end of each preview there was a 2-second pause (R4). The preview was always the first video that was displayed for a task. The mean length of a preview was 28 s (range 21–34).

Demonstrations showed and explained an unfolding scenario of task completion. Research on paper tutorials for software training suggests that the construction of such a scenario revolves around two key design decisions: (a) selecting a suitable solution method and (b) concentrating on the most essential content (Carroll and van der Meij 1998; van der Meij and Carroll 1998).

Software programs generally offer menu-based and key-based methods for task completion. For instance, to italicize a word in a Microsoft Word 2010 document in Windows (it is different for the Mac), the user can select the *I*-object from the Font menu or press the key combination Control + I. The methods have a trade-off between simplicity and efficiency. The recommended choice for novices or beginners is to teach only the simpler, more insightful method (Renkl 2011, 2014; van der Meij and Carroll 1998). In software training this is invariably a menu-based method.

The Four Components Model of procedural discourse for software indicates that tutorial instructions generally consist of four main types of information: goals, interactions, prerequisites, and troubleshooting (van der Meij et al. 2003; van der Meij and Gellevij 2004). Of these types, goals and interactions are a necessary component. Accordingly, mainly information on goals and interactions was given.

Information about the goal serves the dual purpose of motivating and preparing the user. It should identify what a task can accomplish and persuade the user to try to complete that task. Information about the interactions couples the user actions with the program's reactions. The action component must tell the user how to act and what object to act on. These action statements should be aligned to the user's mental plan; they should follow the path of the user's thoughts and actions during task execution (Dixon 1982; Farkas 1999; Zacks and Tversky 2003). The recommended, standard format for an action statement is a succinct description built around an imperative verb (Farkas 1999). The reaction

component should inform the user about the consequences of an action. This information directs the user's attention to interface changes (van der Meij and Gellevij 2004).

Each demonstration began with a display of the starting state of a Word document and a narrative that introduced the formatting problem (e.g., "The left side or margin of this text is too small", "This list does not look nice"). Thereafter, the demonstration showed how user actions led to changes on the interface. The narrative described these actions in command form (e.g., "Click the left mouse button", "Drag the margin to 2.5 cm"). The narrative also regularly attended the user to screen objects, locations and changes using standard phrases such as, "A window appears with the text…" and "You now see…". In addition to this verbal signaling, visual techniques such as zooming-in and highlighting were applied (A1). Statements about actions or interface states frequently included the personal pronoun "you" to create a conversational style (M2). Demonstrations concluded with a reminder of the accomplished goal of the procedure (e.g., "You can now see that all chapter headings look the same. Word also knows exactly where they begin."). At the end of each demonstration there was a 2-second pause (R4). In the 2-demo tutorial, the first demonstration was always followed by a second, identical, task demonstration. The mean length of a demonstration was 53 s (range 42–75).

Reviews (R5) summarized a procedure. They provided the user with a condensed version of task execution. Reviews always appeared automatically after the demonstration for a task, following a 2-second pause (R4). Reviews always began with the announcement "Remember". Thereafter, the narration concentrated on the task sequence and actions. That is, the task completion scenario was emphasized by paying attention mainly to the (sub)goals and user actions needed to achieve them. The narration was expressed in the first person to model the personal action plan, and it used signal words to emphasize the sequential nature of the actions (e.g., "First, I must select a chapter title. Next, from the menu Styles I select Heading 1"). The mean length of a review was 25 s (range 20–35).

A paper *instruction booklet* served to guide the participants through the process of video viewing and practice. It described the sequence in which task videos were to be viewed, and what practice file (see Procedure) to use in the training test.

Instruments

The *initial self-efficacy questionnaire (ISEQ)* was a paper-and-pencil instrument that presented a screenshot of each of the six main training tasks and asks the question: "How well do you think you can complete this task?". The same question was asked for each task that was demonstrated in training. This tailored the measure of self-efficacy to the particular domain of interest (Bandura 2006). Answers were given on a 7-point Likert scale. Following the advice from 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). In presenting the data, the scales are reversed, because that is easier to understand (i.e., a higher score meaning greater self-efficacy). Reliability analysis using Cronbach's alpha indicated satisfactory results for initial self-efficacy ($\alpha = 0.73$).

The *final self-efficacy questionnaire (FSEQ)* was a paper-and-pencil instrument that asked participants to appraise their self-efficacy after training. There were three general questions on ease of learning, and six questions about the formatting tasks (i.e., "Ï can change the left margin of a complete text", "I can change the right margin of a complete text", "I can construct a nice table of contents", "I can present a nicely structured list", "I can indent a citation on the right and left" and "I know how to indent the first line of a new text segment"). Because, the training tasks were familiar, the presentation format no longer

was depicted but merely described. Only the data for the six formatting tasks are reported. Scales and scoring were the same as for the ISEQ. Reliability analysis using Cronbach's alpha indicated good results for final self-efficacy ($\alpha = 0.87$).

Both motivation questionnaires employed in the study had their origin in the Fragebogen Actuelle Motivation (Rheinberg et al. 2001). They were adapted for software training, and used in earlier empirical studies (e.g., van der Meij and van der Meij 2014, 2015).

Five knowledge tests (i.e., *pre-test, training, immediate post-test, delayed post-test,* and *transfer test*) assessed task performance. To ensure that participants understood the pre-test items, task descriptions were coupled to screen shots displaying the start and end state. All tests were performance tests with test tasks that had to be completed on the computer. Test tasks used Word files specially prepared for the experiment. All files, except those in the transfer test, included the same underlying problem demonstrated in the videos, differing only in surface level features. Both the pre-test and training test presented one item at a time. The post-tests presented all six test tasks simultaneously. This made these tests more taxing than the training because participants were now facing the problem of selecting the right method for a specific formatting task (compare Card et al. 1983). The transfer test asked participants to solve four related but untrained tasks (e.g., change the upper margin of a document).

Task performance on the tests was scored with the aid of a codebook. A score of 0 points was awarded for each task the participant did not complete correctly. Correct task completion yielded a score of 1. Separate scores were computed for two training tasks that required two separate subtasks (i.e., citations and list). This led to a maximum score of 8. Scores were converted to a percentage of possible points. Reliability analyses, Cronbach's alpha, led to satisfactory results for all performance tests: pre-test ($\alpha = 0.60$), training ($\alpha = 0.67$), immediate post-test ($\alpha = 0.74$), delayed post-test ($\alpha = 0.73$), and transfer test ($\alpha = 0.76$).

The study used a *logging program* as a proxy for viewing. For every second of video, the program recorded the participant's actions related to the video (i.e., playing, stopping, pausing, or rewinding) and screen setting (i.e., full screen or partial screen display). With the program, 'coverage' could be computed. Coverage was 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 a video. For instance, when a participant's log revealed that there were 26 s of unique play events for a 28-s preview, the coverage for that preview was computed to be 93% (26/28). The minimum score of 0% indicated that a video was never set into play mode. The maximum score of 100% indicated that every unique video second had been played at least once. Coverage was computed for all videos together (total coverage) and for videos types (i.e., previews, 1st demonstrations, 2nd demonstrations, and reviews).

Procedure

The study was conducted in three sessions held in the computer room at the school. All sessions were led by the researcher. In the first session, participants were told that they would engage in software training on Word to assist them in improving the formatting of their school reports (5 min). Next, they were instructed to complete the ISEQ (5 min) and pre-test (20 min).

Training followed a week later. This session started with a 10-min introduction that explained Website navigation and video usage. Also the use of the instruction booklet was

explained. Participants were instructed to view the task video until they felt sure they could achieve the procedure. Thereafter, they could engage in practice on the task but they were not allowed to look back to the video. The training included periods of practice (P1) because production processes could not be directly supported by the videos themselves (see Fig. 1). Practice on a task immediately followed the video for that task. This timing of practice (P2) is beneficial for task execution during training. Practice was further supported by practice files (P3) which kept distracting information to a minimum. Participants were told to use headphones and work independently, and to call for assistance only when needed. They were given 45 min of training time in the control condition and 50 min in the experimental conditions to compensate for longer videos. [With one exception, this was sufficient for all participants.] After training, there was a 10-min break, followed by the FSEQ and the immediate post-test (20 min). Participants were not allowed to consult the videos during testing.

The third session followed 1 week later. Participants first worked on the delayed post-test (20 min). Next, the participants were informed about the transfer test, in which they were to complete four new, untrained tasks (5 min). Then the transfer test was taken (20 min).

Analysis

A check on the random distribution of participants across conditions revealed no significant differences for age, $F(2,62) = 0.28$, $p = 0.973$, or gender, $\chi^2(2,65) = 0.94$, $p = 0.625$. Conditions also did not differ on initial self-efficacy, $F(2,61) = 0.14$, $p = 0.873$, or on pre-test scores, $F(2,64) < 1$, n.s. There was also no effect of condition on total training time, $F(2,64) < 1$, n.s. The data set from two participants (one in the control conditions, and one in the 2-demo condition) was incomplete. These participants were included in the analyses whenever possible. AN(C)OVAs were computed to assess differences in coverage, motivation and learning. For significant main results, contrasts were computed using Tukey's least significant difference (LSD). Initial checks for sphericity violations (Mauchly) or assumptions of homogeneity of variance (Levene test) revealed no significant differences. Tests were two-tailed with alpha set at 0.05. The degrees of freedom occasionally differ due to missing data. Cohen's (1988) d-statistic was used to report effect size. These tend to be characterized as small for $d = 0.2$, medium for $d = 0.5$, and large for $d = 0.8$.

Results

Video coverage

The mean total coverage was 70.2% (see Table 2). An ANOVA showed a main effect of condition, $F(2,62) = 6.49$, $p = 0.003$. Post hoc analyses revealed that total coverage for the 2-demo tutorial was lower than for the control ($p = 0.001$), and the review tutorial ($p = 0.041$). The control and review tutorial did not differ for total coverage ($p = 0.149$).

A repeated measures ANOVA with condition as independent variable, and previews and first demonstrations as the two measurement points, showed a significant decline over time, $F(1,62) = 76.8$, $p < 0.001$, $d = 1.21$. Coverage was lower for first demonstrations than for

Table 2 Means (standard deviations) for coverage by condition and video type

Condition	Total		Preview		1st demo		Review		2nd demo	
	M	(SD)	M	(SD)	M	(SD)	M	(SD)	M	(SD)
Control (n = 22)	79.8	(22.6)	92.0	(18.8)	70.2	(27.8)	n.a.		n.a.	
Review (n = 21)	71.4	(16.0)	96.3	(7.5)	72.1	(22.0)	42.5	(30.0)	n.a.	
2-Demo (n = 22)	59.3	(17.4)	95.4	(9.8)	72.0	(21.5)	n.a.		27.5	(22.2)
Total (n = 65)	70.2	(20.5)	94.5	(13.0)	71.4	(23.6)	42.5	(30.0)	27.5	(22.2)

previews. There was no main effect of condition, $F(2,62) < 1$, and no interaction, $F(2,62) < 1$.

A repeated measures ANOVA comparing the two experimental conditions, with review or second demonstration as a third time point, showed a significant main effect of the factor of time, $F(2,82) = 167.4$, $p < 0.001$. More importantly, there was a significant interaction between condition and time, $F(2,82) = 3.1$, $p = 0.05$.

An ANCOVA comparing reviews and second demonstrations, with preview and first demonstration coverage as covariates, indicated a significant effect of condition, $F(1,39) = 6.7$, $p = 0.013$, $d = 0.57$. Table 2 shows that reviews yielded moderately higher video coverage than second demonstrations.

Exploratory analyses of the factor of time (the sequence of nine task videos) for coverage per video type revealed that previews yielded consistently high scores (range 89.8–97.7). Other video types displayed significant declines over time. The coverage for demonstrations, reviews, and second demonstrations all dropped substantially as participants continued on in the video tutorial.

Self-efficacy

The initial scores for self-efficacy were slightly above the scale mean (see Table 3). A repeated measures ANOVA with condition as independent variable and initial and final self-efficacy reports as the two measurement points, showed a main effect of time. The video tutorials enhanced self-efficacy significantly and substantially, $F(1,61) = 90.02$, $p < 0.001$, $d = 1.48$. There was no main effect of condition and no interaction between time and condition, $F < 1$.

Table 3 Means (standard deviations) for self-efficacy (mean scale value is 4; scale maximum is 7. A higher score indicates stronger self-efficacy) by condition

Condition	Self-efficacy			
	Before		After	
	M	(SD)	M	(SD)
Control (n = 21)	4.33	(1.51)	6.07	(1.19)
Review (n = 21)	4.22	(1.34)	6.33	(0.55)
2-Demo (n = 22)	4.43	(1.18)	6.05	(1.40)
Total (n = 64)	4.33	(1.33)	6.11	(1.11)

Learning

Table 4 shows the results for the test outcomes. The initial level of performance was low, with a mean score of 11.7%. A repeated measures ANOVA with condition as the independent variable and the regular task tests as four measurement points, showed a main effect of time, $F(3,183) = 202.74$, $p < 0.001$. Post hoc analyses revealed a significant change from pre-test to training ($p < 0.000$) with scores reaching an average of 79.7% (see Table 4). There was also a significant change from training to immediate post-test ($p < 0.001$), with scores declining to 62.5%. Scores on the immediate and delayed post-tests did not differ ($p = 0.43$). Also, there was no main effect of condition and no interaction between time and condition, $F < 1$. Furthermore, an ANCOVA with condition as the independent variable, pre-test score as the covariate and the transfer test as the dependent variable, revealed no difference between conditions, $F < 1$.

Discussion

The findings for coverage, a proxy for viewing, yielded an average of 70%. In other words, more than two-thirds of each task video in the tutorials was shown in play mode and, therefore, had the potential of being viewed by the participants. Video type affected this outcome. Previews yielded coverage scores of well over 90%, while coverage dropped significantly for the later parts of each task video. First demonstrations scored about 70%; reviews and second demonstrations yielded scores of 42 and 27%, respectively.

One factor that appears to have had a negative effect on coverage is placement in the tutorial. Whereas individual task videos stayed well within the 3-min limit recommended by Guo et al. (2014), the total for the whole set of videos in the tutorial was considerably longer, namely 12–20 min in the case of complete coverage (see "Instructional materials" under "Method"). Exploratory analyses indicated that there was a decrease in coverage for first demonstrations, second demonstrations, and reviews as participants got further along in the tutorial.

Coverage may also have been negatively affected by redundancy. Although we tried to optimize the design of each video type for the purpose(s) that it served, some overlap between types for the same task was inevitable. The steep decline that occurred from previews to first demonstration videos is perhaps a signal that participants found that the latter video did not contain enough new information for them to keep viewing. They may

Table 4 Means (standard deviations) for test scores by condition and test

Condition	Pre-test		Training		Immediate post-test		Delayed post-test		Transfer test	
	M	(SD)	M	(SD)	M	(SD)	M	(SD)	M	(SD)
Control (n = 22)	9.7	(12.8)	76.1	(18.9)	58.5	(26.6)	55.1	(25.2)	33.0	(34.0)
Review (n = 21)	11.3	(15.3)	79.8	(22.5)	67.2	(26.9)	68.5	(25.8)	29.8	(36.8)
2-Demo (n = 21)	14.3	(19.1)	83.3	(19.5)	61.9	(29.7)	58.3	(25.7)	29.8	(35.0)
Total (n = 64)	11.7	(15.8)	79.7	(20.2)	62.5	(27.5)	60.6	(25.8)	30.9	(34.7)

have decided to drop out earlier with first demonstrations because they believed they had already seen enough of what was needed for task execution in the preview.

Still, the distinction between preview, demonstration, and review videos is valuable because each video type supports a different process in observational learning. This raises the question how can videos be redesigned to yield higher coverage?

One possibility is to reduce the content of the previews. Currently, the previews prepare participants cognitively and motivationally for the forthcoming demonstration by informing them about the goal, the concept and the interface object(s) involved in task achievement. Thus, the structure of previews may have given participants the false impression that they presented a comprehensive view on task accomplishment. Perhaps previews should concentrate on explaining only the concepts, just as in designs for the pre-training principle (see Mayer and Pilegard 2014).

Another possible design solution is to use different examples. All videos for a task featured the same sample file. This design choice was based on the consideration that an identical example throughout a task case would free up cognitive space for attending to other, more pertinent task features because familiar task aspects are more easily ignored (Schnotz and Bannert 1999). Apparently, it did not work that way. Rather the presentation of the same example may have given participants the impression that the content was already known, which worked against continued viewing of all task videos. Akin to the desirable differences hypothesis (see Yue et al. 2013), it is probably better to aim for a design that makes participants perceive all video types for a task as slightly different.

One goal of the video tutorials was to provide motivational support to users. The findings showed that the tutorials contributed significantly and substantially to self-efficacy development. Participants indicated that the tutorials resulted in high final self-appraisals of the capacity to accomplish the trained tasks in the future.

The gain in self-efficacy is relevant for future actions, as high self-efficacy has been found to be related to trying harder and being more persistent with comparable task challenges (Bandura 2012; Bandura and Locke 2003). More generally, the outcome is relevant in view of the prominent place given to motivation in the theoretical model (see Fig. 1), along with the motivational mediation assumption, which holds that motivational factors mediate learning by increasing or decreasing cognitive engagement (Park et al. 2014).

Future studies might want to obtain a broader picture of the motivating qualities of video tutorials by gathering the students' opinion about aspects such as ease of use and usefulness. The technology adoption model (e.g., Davis 1989; Davis et al. 1989; Joo et al. 2014) suggests that these appraisals can play an important role in the users' willingness to engage with video.

Another goal of the video tutorials was the provision of cognitive support. The data showed that the tutorials effectively supported learning when assessment followed immediately after the task videos. From a meager initial score of 12% of the items successfully completed on the pre-test, the training test items yielded a score of 80%. Later testing points revealed that there were still significant and substantial learning gains compared to the starting level, but also that successful completion dropped to about 60%.

Compared to our first empirical study on video tutorials with and without reviews (van der Meij and van der Meij 2016a, b), the learning outcomes are substantially lower on both the immediate and delayed post-test, and the superiority of the review tutorial also failed to replicate. An important difference with the present study lies in a situational variable (i.e., procedure). In the first experiment, participants could access the videos at all times during

training. In the present study, access was prohibited when participants worked on practice tasks.

Why is the former more beneficial for learning? According to Bandura (1986), modeling and production should be interleaved. One reason why participants need immediate practice after observing a model is that it reveals where their imitation of the model fails. When they cannot yet do what the model showed it can prompt participants to replay video segments. Practice thus provides participants with a litmus test for how well a procedure for task completion is understood and remembered, and also serves as a cue for where retention failed. This, in turn, should be a strong stimulus for replaying (part of) the video. In addition to addressing any gaps, replays can solidify the knowledge that has already been acquired. The latter would appear to be especially important when subsequent videos discuss closely related procedures.

The participants in the present study did well during practice (i.e., on the training test), but their scores on the immediate and delayed post-test suggest that learning was fragile. Blocking access to the videos may have affected consolidation. We had expected the reviews to compensate for the lack of video access. The data showed that the scores on the immediate and delayed post-test were about 10% higher than in the other conditions. However, the difference was not statistically significant. The meager coverage for reviews probably limited their impact.

Conclusion

The present study was conducted to further investigate the effectiveness of video tutorials for software training that are designed according to a theoretical model that combines insights from DBT (Grossman et al. 2013) and multimedia learning theory (e.g., Mayer 2014). The findings showed that the tutorials successfully supported and enhanced motivation, but also that a lasting effect on learning was achieved only to a moderate degree. One factor that probably negatively affected learning was the restriction imposed on practice; this belongs with the 'situational variables' in the theoretical model. By blocking access to the video during training tasks, the participants were limited in their opportunity for learning.

In addition, the study investigated whether a video tutorial would be more effective if it included reviews. The expected benefits for reviews were not found, which was ascribed to their low coverage. One explanation that was advanced for the limited viewing of reviews was tutorial length. Another factor that could have played a role concerned the content of the reviews. Future studies might want to explore whether reviews achieve higher viewing rates when the modeled task involves a (superficially) different example than in the demonstration. Alternatively, it seems important to assess whether participants value reviews. Because reviews are not commonly presented at the end of instructional videos it seems important to gather the participants' opinions about the usefulness of reviews.

Future research should also address the limitation that no process information was gathered. Empirical research on signaling warns that effective support for a single observational learning process (i.e., attention) does not necessarily yield greater learning (de Koning et al. 2010; Kriz and Hegarty 2007; Skuballa et al. 2012). In other words, this research suggests that it is desirable to obtain process measures for other activities involved in observational learning as well. For instance, a think-aloud protocol during the practice

tasks or the post-tests might give information about the retention process when information from the narrative is sought for, repeated, or summarized by the participants.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval A research proposal describing the study has been submitted to the ethics committee of the University who has given approval.

References

Alexander, K. P. (2013). The usability of print and online video instructions. *Technical Communication Quarterly, 22,* 237–259. https://doi.org/10.1080/10572252.2013.775628.

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: Freeman and Company.

Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares & T. C. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (pp. 307–337). Greenwich: Information Age Publisher.

Bandura, A. (2012). On the functional properties of perceived self-efficacy revisited. *Journal of Management, 38*(1), 9–44. https://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. https://doi.org/10.1037/0021-9010.88.1.87.

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

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

Carroll, J. M., & van der Meij, H. (1998). Ten misconceptions about minimalism. In J. M. Carroll (Ed.), *Minimalism beyond the Nurnberg Funnel* (pp. 55–90). Cambridge, MA: MIT Press.

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

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340.

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science, 35*(8), 982–1003.

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. https://doi.org/10.1016/j.compedu.2010.02.027.

Dixon, P. (1982). Plans and written directions for complex tasks. *Journal of Verbal Learning and Behavior, 21,* 70–84. https://doi.org/10.1016/S0022-5371(82)90456-X.

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

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, DC.

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. https://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.* Paper presented at the L@S '14, March 4–5, 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. https://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. https://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.

Joo, Y. J., Lee, H. W., & Ham, Y. (2014). Integrating user interface and personal innovativeness into the TAM for mobile learning in Cyber University. *Journal of Computing in Higher Education, 26*(2), 143–158. https://doi.org/10.1007/s12528-014-9081-2.

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

Kalyuga, S. (2014). The expertise reversal principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 576–597). New York: Cambridge University Press.

Kosoff, M. (2017). LinkedIn just bought online learning company Lynda for $1.5 billion. Retrieved May 2, 2017 from http://www.businessinsider.com/linkedin-buys-lyndacom-for-15-billion-2015-4?international=true&r=US&IR=T.

Kriz, S., & Hegarty, M. (2007). Top-down and bottom-up influences on learning from animations. *International Journal of Human Computer Studies, 65,* 911–930. https://doi.org/10.1016/j.ijhcs.2007.06.005.

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

Mayer, R. E. (2008). *Learning and instruction* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

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

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: Cambridge University Press.

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

Mestre, L. S. (2012). Student preference for tutorial design: A usability study. *Reference Service Review, 40*(2), 258–276. https://doi.org/10.1108/00907321211228318.

Meyer, B. J. F. (1975). *The organization of prose and its effects on memory*. Amsterdam: North Holland.

Meyer, B. J. F. (1984). Text dimensions and cognitive psychology. In H. Mandl, N. Stein, & T. Trabasso (Eds.), *Learning and understanding texts* (pp. 3–51). Hillsdale, NJ: Erlbaum.

Meyer, B. J. F., & Poon, L. W. (2001). Effects of structure strategy training and signaling on recall of text. *Journal of Educational Psychology, 93,* 141–159.

Meyer, B. J. F., & Wijekumar, K. (2007). A web-based tutoring system for the structure strategy: Theoretical background, design, and findings. In D. S. McNamara (Ed.), *Reading comprehension strategies. Theories, interventions, and technologies* (pp. 347–374). New York: Lawrence Erlbaum Associates.

Park, B., Plass, J. L., & Brinken, R. (2014). Cognitive and affective processes in multimedia learning. *Learning and Instruction, 29,* 125–127. https://doi.org/10.1016/j.learninstruc.2013.05.005.

Renkl, A. (2011). Instruction based on examples. In R. E. Mayer & P. A. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 272–295). New York: Routledge.

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

Rheinberg, F., Vollmeyer, R., & Burns, B. D. (2001). FAM: Ein Fragebogen zur Erfassung aktueller Motivation in Lern- und Leistungssituationen. *Diagnostica, 47*(2), 57–66. https://doi.org/10.1026//0012-1924.47.2.57.

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. https://doi.org/10.1177/0018720810381071.

Schnotz, W., & Bannert, M. (1999). Einflüsse der visualisierungsform auf die konstruktion mentaler modelle beim text- und bildverstehen. *Zeitschrift für Experimentelle Psychologie, 46,* 217–236. https://doi.org/10.1026//0949-3964.46.3.217.

Skuballa, I. T., Schwonke, R., & Renkl, A. (2012). Learning from narrated animations with different support procedures: Working memory capacity matters. *Applied Cognitive Psychology, 26,* 840–847. https://doi.org/10.1002/acp.2884.

TechSmith. (2017). SnagIt tutorials. Retrieved June 4, 2017 from https://www.techsmith.com/tutorial-snagit-13.html.

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

van der Meij, H., Blijleven, P., & Jansen, L. (2003). What makes up a procedure? In M. J. Albers & B. Mazur (Eds.), *Content & complexity. Information design in technical communication* (pp. 129–186). Mahwah, NJ: Erlbaum.

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., & Gellevij, M. R. M. (2004). The four components of a procedure. *IEEE Transactions on Professional Communication, 47*(1), 5–14. https://doi.org/10.1109/TPC.2004.824292.

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*(2), 116–132.

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

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

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.

Wistia. (2012). Does length matter? Retrieved May 7, 2012 from http://wistia.com/blog/does-length-matter-it-does-for-video-2k12-edition.

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*(2), 266–277. https://doi.org/10.1037/a0031971.

Zacks, J. M., & Tversky, B. (2003). Structuring information interfaces for procedural learning. *Journal of Experimental Psychology: Applied, 9*(2), 88–100. https://doi.org/10.1037/1076-898X.9.2.88.

Hans van der Meij is Senior Researcher and Lecturer in Instructional Technology at the University of Twente (The Netherlands). His research interests are: technical documentation (e.g., instructional design, minimalism, the development of self-study materials), and the functional integration of ICT in education. He has received several awards for his articles, including a "Landmark Paper" award from IEEE for a publication on minimalism (with John Carroll).

Jan van der Meij is Assistant Professor at the Department of ELAN Teacher Development and Director of the teacher training institute of the University of Twente. He received his Ph.D. in 2007. His research concentrates on functional use of ICT in education. This includes learning with (multiple) dynamic representations, (live) video instruction, and video feedback coaching.

Tessa Voerman completed her M.A. study in the Learning Sciences at the University of Twente. She is currently pursuing a M.A. study in Educational Science and Technology at the University of Twente.

Evert Duipmans completed his M.A. study in informatics at the University of Twente. He is currently employed as a teacher at Saxion.