Computers in Human Behavior 89 (2018) 439–445

Contents lists available at ScienceDirect

Computers in Human Behavior

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

Effects of practice with videos for software training

Hans van der Meij [a,*], Ilona Rensink [b], Jan van der Meij [c]

[a] University of Twente, Faculty of Behavioural, Management and Social Sciences, Educational Science and Technology, P.O. Box 217, 7500 AE Enschede, The Netherlands
[b] Diepenveenseweg 124, 7413 AT Deventer, The Netherlands
[c] University of Twente, Faculty of Behavioural, Management and Social Sciences, ELAN Teacher Development, P.O. Box 217, 7500 AE Enschede, The Netherlands

ARTICLE INFO

Article history:
Received 27 February 2017
Received in revised form
9 October 2017
Accepted 20 November 2017
Available online 21 November 2017

ABSTRACT

This study investigated the contribution of practice in learning from Demonstration-Based Training (DBT) videos for software training. An experiment with three conditions is reported: practice followed by video (practice-video), video followed by practice (videopractice), and video only. The combination of practice and video was expected to enhance learning more than the video only condition. Also, practice-after was expected to be more effective than practice-before. The 82 participants, elementary students (mean age 11.2), achieved significant learning gains, reaching moderate to high levels of success on the immediate and delayed post-tests, and the transfer test. No practice effect was found. Also, there was no difference in test performance between practice conditions. The discussion advances several options for enhancing the effectiveness of the DBT-videos.

1. Introduction

Instructional videos (videos, from here on) for software training are becoming commonplace. Software companies such as Adobe, IBM, Microsoft, SAP, and TechSmith are offering more and more videos on their websites. These videos usually consist of a recorded demonstration — a screen capture animation with narration. Beyond that, though, little is known about the design characteristics and effectiveness of the videos produced by these companies. When using such videos to provide a tutorial for an audience of novices, the videos must generally accomplish two goals. One objective is to support task performance; the videos must enable or guide the user's task completion. Their other role is to support learning; the videos must instruct the user so that he or she can acquire the capability to perform trained and related tasks independently (Grabler, Agrawala, Li, Dontcheva, & Igarashi, 2009; van der Meij, Karreman, & Steehouder, 2009).

Design of videos should be oriented toward achievement of both of these goals. Recent research on software training has proposed and tested a theoretical model for video construction that combines Demonstration-Based Training (DBT) and multimedia learning theory (e.g., Brar & van der Meij, 2017; van der Meij & van der Meij, 2016a; van der Meij, 2017). The videos in these studies illustrate and explain the stepwise progression involved in task completion. That is, each video shows a single, menu-based method for completing the given task. That demonstration is enhanced with instructional features that support four key observational learning processes, namely, motivation, attention, retention, and (re)production. The inclusion of practice to support the (re)production process is important for the goal of learning, but its contribution has rarely been empirically investigated. It is the focus of this paper.

The present study investigates the influence of the presence and timing of practice on task completion and learning in video-based software training. Because very few studies have investigated the inclusion of practice in such training, we begin with a review of the research on practice in the related field of worked examples and then review practice in videos for software training. After that, we introduce and report an experiment with varying practice conditions in video-based software training.

2. Worked-examples research and the presence and timing of practice

The design of DBT-based videos for software training bears great similarity to the design of worked examples, which have a long and

* Corresponding author.
E-mail address: H.vanderMeij@utwente.nl (H. van der Meij).

https://doi.org/10.1016/j.chb.2017.11.029

successful tradition in the development of problem-solving skill, usually in the domain of science (e.g., Atkinson, Derry, Renkl, & Wortham, 2000; Renkl, 2014; Salden, Koedinger, Renkl, Aleven, & McLaren, 2010; Sweller & Cooper, 1985; van Gog & Rummel, 2010). A worked example draws students' attention to key features in a problem and provides them with domain- or task-specific information. In other words, a worked example presents an ideal model that gives a step-by-step explanation of problem solving.

Many worked-example designs have a classic coupling of instruction and practice, with the former preceding the latter; students first receive procedural instructions (the worked example) about a topic and then engage in practice on a similar problem. In worked-examples research, the contribution of practice to learning has become a topic of systematic investigation (e.g., Leppink, Paas, van Gog, van der Vleuten, & van Merriënboer, 2014; Reisslein, Atkinson, Seeling, & Reisslein, 2006; van Gog, Kester, & Paas, 2011).

Worked examples research has advanced several arguments for including *practice-after giving instructions*. One reason is that practice-after can deepen understanding. Exercises, problems, or tasks given after the worked example stimulate students to construct meaning. A risk of worked examples is that they offer convenient directions that may invite passive and superficial processing (Atkinson et al., 2000). When students do not reflect on the examples sufficiently, the effectiveness of the examples is seriously threatened. The inclusion of practice-after can stimulate such reflective activities.

Another reason is that practice-after can consolidate learning. After having seen the modelled performance, practice serves as a check of understanding. The worked example provides students with a mental model of the solution process which can then be consolidated by practicing with a similar problem (van Gog, 2011).

Empirical studies have generally supported the claim that practice-after effectively increases learning of novices. Reisslein et al. (2006) found that low prior knowledge participants did better with practice-after and high prior knowledge participants did better with practice-before for worked examples on problem solving in electric circuits. Wouters, Paas, and van Merriënboer (2010) examined the role of practice of trained and transfer tasks with an animated model for problem solving in probability calculus. They compared practice-after, practice-before, and restudy of the worked examples and found no differences between conditions. The absence of the predicted advantage of the practice-after condition was explained by the fact that the participants in the study had relatively high prior knowledge. van Gog et al. (2011) compared example only, practice only, example with practice-after, and example with practice-before for four electrical circuit troubleshooting tasks. The findings on an immediate post-test showed significantly higher scores for the example only and practice-after conditions than for the practice only or practice-before conditions. No difference was found between the example only and practice-after conditions. Leppink et al. (experiment 2, 2014) replicated these findings for two application tasks on Bayes' theorem in statistics.

The prevalent argument for including *practice-before instructions* concerns a certain condition for learning, namely, the level of prior knowledge. When students have high prior knowledge, they may benefit more from an opportunity for exploration or orientation before receiving instructions. One reason is that these activities stimulate students to make connections between known and new information (Kalyuga, 2007).

Another argument in favor of practice-before instructions is that such a sequence is more motivating. Practice stimulates students to think hard about a problem. In turn, this also motivates them to process the subsequent worked example more deeply (Stark, Gruber, Renkl, & Mandl, 2000).

Only a few empirical studies have investigated the relative effectiveness of practice-before instruction. The studies by Reisslein et al. (2006) and Wouters et al. (2010) were discussed earlier. In addition, Paas (1992) found that participants in a worked examples only condition had significantly higher scores compared to practice-before on trained and transfer items concerning problem solving in statistics. In a study on Computer-Numerically-Controlled machinery, Paas and Van Merriënboer (1994) also found higher transfer test scores for the examples only condition. In contrast, Stark et al. (2000) predicted and found a stimulating effect of practice-before. Their experiment on computing interest rates showed a carry-over effect from one to the next of five successive practice-example pairs.

Overall, the findings from systematic research on the role of timing of practice in worked examples have been slightly mixed. Most studies with novices have found that practice-after is more effective for learning than practice-before, the exception being the study by Stark et al. (2000). Several studies have also found that example only study was equally as effective as practice-after.

3. Practice in videos for software training

To our knowledge, only Ertelt (2007) has studied the role of practice in video-based software training. We have therefore also included in our review of the empirical literature two (older) studies that we found on video-based instructions for assembly tasks, because such instructions set out to achieve a similar type of learning outcome as in software training.

Hannafin and Colamaio (1987) examined the influence of *practice-after* instructions for a 30-min video on a resuscitation apparatus. They found no improvement in task performance from the inclusion of practice-after. To account for the absence of an effect of practice, the authors postulated that participants engaged in a “form of vicarious mental rehearsal” that made “overt practice in the procedure unnecessary” (p. 210). Baggett (1988) also found no effect of practice-after. Her study compared a video-only with a video with practice-after condition for a set of Fischer-Technik assembly instructions. She suggested that the video alone sufficiently facilitated learning by enabling the participants to “imagine the (correct) movement” (p. 496). Ertelt (experiment 2, 2007) found a significant but small positive effect of practice-after on learning from a video for software training. It was suggested that the inclusion of practice had a motivating effect on the students.

We were unable to find any experiments involving video-based software training that manipulated the presence of *practice-before*. The absence of such studies is somewhat surprising considering the fact that users frequently consult software videos after having tried and failed to complete a software task (van der Meij et al., 2009).

4. Experimental design and research questions

The present study was set up as a quasi-experimental design with random allocation of participants within classrooms to conditions. There were three conditions: (1) Practice-video, (2) Video-practice, and (3) Video. The first two conditions combined watching the video with hands-on practice, the only difference being their sequencing. Participants in the practice-video condition first tried to complete a practice task themselves before receiving video instructions for that task, whereas participants in the video-practice condition followed the reverse order. Participants in the third condition only watched the videos. Learning was assessed with performance tests that required completion of tasks that differed from the demonstrated tasks in superficial features only.

Information on personal characteristics (e.g., gender, age, and prior task experience) was collected before training. The participants' motivation was also measured prior to training to check on random distribution across conditions, because motivation plays a key role in DBT-based video design, and may affect task performance and learning from instructional video. For motivation, the two key constructs from expectancy-value theory (Eccles & Wigfield, 2002) were measured. For value, we measured task relevance, which refers to the perceived present and future utility of an activity. It is indicative of the importance of a task to a person's goals or concerns (van der Meij, 2007). Self-efficacy can be defined as a person's expectancy to succeed in specific situations or accomplish a task (Bandura, 1997). When self-efficacy is high, people are more likely to attempt new tasks and to persist when obstacles occur.

Research question 1: "Does practice enhance learning from a video tutorial?" Practice calls on the user to engage in reproductive activities, which Bandura (1986) social-cognitive learning theory describes as an important process in observational learning. In addition, because there is probably a motivating and thought-provoking effect of engaging in practice (e.g., Atkinson et al., 2000; Stark et al., 2000), it is hypothesized that learning will be higher in the two practice conditions.

Research question 2: "Is there an effect of timing of practice?"

The standard procedure in worked-examples research is to couple instructions and practice, with the former preceding the latter. Likewise, a consistent recommendation in training for procedural knowledge development is to include practice after the instruction (Ertelt, 2007; Grossman, Salas, Pavlas, & Rosen, 2013; Kerr & Payne, 1994; Rosen et al., 2010; van der Meij & van der Meij, 2013). Furthermore, worked-example research has posited a (more) positive effect of practice-after on the grounds that it enhances understanding and consolidates learning (Atkinson et al., 2000; van Gog, 2011). Therefore, the tested hypothesis is that the highest learning outcomes will be found in the condition in which video is followed by rather than preceded by practice.

5. Method

5.1. Participants

The participants consisted of 82 students from two fifth- and two sixth-grade classrooms from three elementary schools in the Netherlands. All schools provided consent for participation of their students. There were 40 girls and 42 boys (mean age 11.2 years; range 10.0–13.0). Within each classroom, students were randomly split into three groups that were randomly assigned to conditions, yielding 27 students in the practice-video, 28 students in the video-practice, and 27 students in the video condition. A check on their distribution revealed no differences between conditions for age or gender.

5.2. Instruments

Videos. The videos concerned several Microsoft Word formatting tasks that are important for the school reports that the students need to create. They were organized into three 'chapters'. The first dealt with adjusting the left and right margins for an entire document (2 videos). The second concentrated on formatting paragraphs, citations, and lists (4 videos). The third dealt with automatically generating a table of contents (3 videos). Earlier studies had indicated that students from this age group generally do not yet know the best method, if any, for accomplishing these tasks (van der Meij, 2014, 2017; van der Meij & van der Meij, 2016a, b).

Each video revolved around a dynamic demonstration of task performance that was accompanied by a female voice-over. Thus, each video displayed and explained an unfolding scenario of task completion. Videos presented only a single, menu-based method for accomplishing the task (see e.g., Carroll & van der Meij, 1998; Renkl, 2014). Conceptual explanations were given strictly on a need-to-know basis. The average length of the videos was 1.13 min (range 0.47–1.42).

Students could access the videos via a website that presented a table of contents in which the chapter titles served as organizers. The table was permanently visible to ensure access to the videos (see Fig. 1). Paragraph titles described the main tasks for each chapter, and linked to the videos. Clicking on a paragraph title changed its color, and opened the video on the right-hand side of the screen. Students could play and pause the video with a control panel.

Instruction booklet. An instruction booklet guided the students through training, providing a scenario of reading, viewing (and doing) for students to follow. The booklet had the same domain structure displayed in the table of contents on the website. For new tasks, the booklet gave an explanation of the concept involved (e.g., margin, paragraph). In addition, there were screenshots of the initial and final screens for tasks to illustrate the goals of the instruction and to motivate the students to try to attempt them. The booklet presented this information as a "Reading task" (see the top part of Fig. 2). Next, the booklet instructed students to watch the video, and possibly also to engage in hands-on practice. The "Viewing task" and "Practice task" were presented as two additional, distinct information types in the booklet (see the bottom part of Fig. 2). The order in which the students were instructed to view the video and engage in practice varied for the two practice conditions in the experiment.

Initial Experience and Motivation Questionnaire (IEMQ). This questionnaire measured the students' initial motivation. For each formatting task, the student first received a Before-After screenshot plus explanation, and was then asked three questions: (a) "Do you ever experience this situation?" (Task experience), (b) "How often do you need to complete this task?" (Task relevance), and (c) "How well do you think you can complete this task?" (Self-efficacy). The student answered these questions by circling a number on a 7-point Likert scale where the end points were given as *never – always*, or *very poorly – very well*. Good reliability scores (Cronbach alpha) were found for task experience (0.78), task relevance (0.85), and self-efficacy (0.81). A mean score was computed for the three constructs in the IEMQ.

Tests. Learning was assessed with performance tests (i.e., pre-test, practice, immediate and delayed post-test, transfer test). Each test, except the transfer test, asked the student to complete formatting tasks demonstrated in the videos (i.e., adjusting the right and left margin of a document, indenting a citation to the left and right, indenting the first line of the paragraphs in a document, improving the presentation of a list, and creating an automatic table of contents). To complete the pre-test, practice (if any),

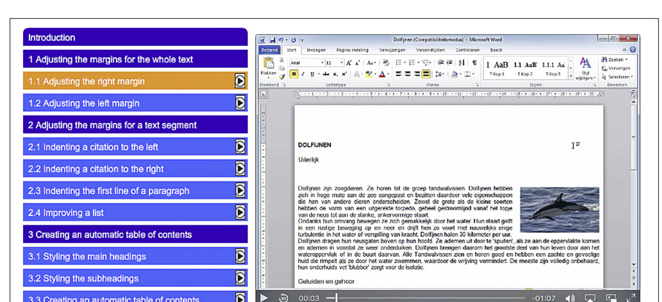

Fig. 1. Screenshot from the video tutorial website (original was in Dutch).

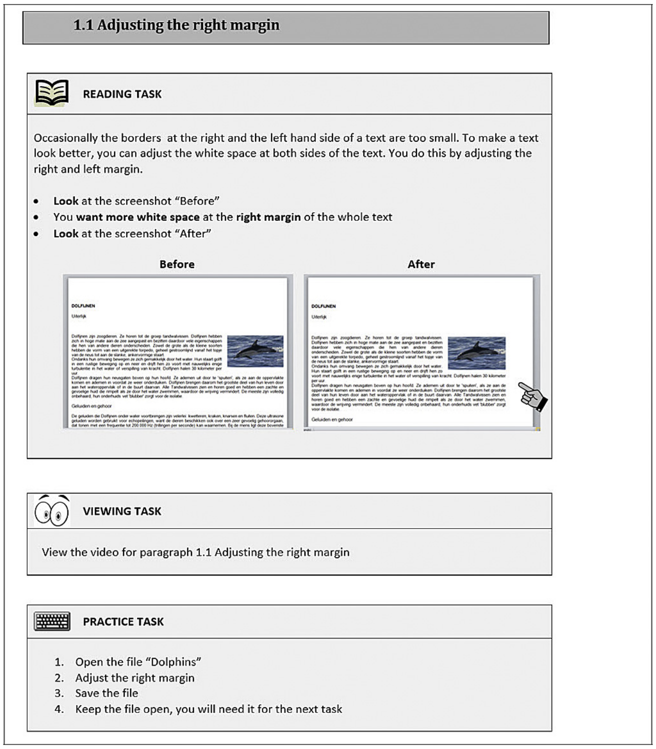

Fig. 2. A page from the instruction booklet (original was in Dutch).

immediate, and delayed post-tests, participants were presented with Word files that were superficially different but structurally identical to the demonstrated tasks. These similar files enabled students to concentrate on the formatting problem (see van der Meij & Carroll, 1998). The transfer test also provided participants with a dedicated Word file. This test involved formatting tasks that varied slightly from the demonstrated tasks. The transfer test consisted of three items, one for each 'chapter'. Item one required a change in the top and bottom margins of an entire text. Item two involved a 3-cm indentation of the items on the second level of a list. Item three required movement of a specific paragraph in a document and an update of the table of contents.

Scoring was the same for all tests. That is, a score of 1 point was awarded for each task that the student formatted correctly. Incorrect or incomplete task performance received a score of 0. Scores were converted to a percentage of possible points. Satisfactory reliability scores (Cronbach alpha) were found for the pre-test (0.68), practice (0.81), immediate post-test (0.71), and delayed post-test (0.76). The reliability score for the transfer test was low (0.47), which may be due to the varied nature of these test items. In view of the poor reliability, the findings for this test are not reported.

5.3. Procedure

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

The training session followed a day later. This session started with a 10-min introduction. An explanation was given about using the instruction booklet. This explanation told the students about the different types of activities (and their sequence) in which they were expected to engage. Website navigation and video usage were also explained and illustrated. In all conditions, the students were instructed to watch the video until they felt sure they could complete the task. Students in both practice conditions received further instructions about the handling of practice files. In the practice-video condition, these students were also told that they were not allowed to return to the practice task after having seen the video, but should continue to the next practice task. In the video-practice condition, the students were told that they were not allowed to return to the video once they started to practice on the just-viewed tasks. The students were instructed to work independently for 50 min and to call for assistance only when stuck. Students received the audio input from the video via headphones. After training was completed, there was a 10-min break followed by the immediate post-test, for which the students were given 20 min. Students were not allowed to consult the video during this (or the delayed) test.

The third session followed one week later. In a brief (5 min) introduction the students were told that, in addition to a test on the trained tasks, there was another test with three new, untrained tasks that they were to try to complete. The students had 30 min to complete both tests.

5.4. Analysis

A non-parametric (i.e., Kruskall-Wallis) test was employed to check on the random distribution of the IEQM data across conditions, because data explorations revealed violations of the normality assumption. Repeated testing was taken into account by applying a Bonferroni correction. That is, with three tests, alpha was set at 0.017 (two-tailed).

The main findings were also tested non-parametrically, because here too data explorations revealed non-normal distributions (e.g, skewed left for the pre-test and skewed right for practice). First, a Friedman analysis was conducted on the four performance measurements. This test showed a statistically significant overall effect of time, $F = 99.02$, $p < 0.001$. Next, Wilcoxon's T-tests were used to assess differences between two moments of measurement. For these outcomes, the z-scores will be reported. Kruskall-Wallis's H-test was used to assess the effects of condition (the Mann-Whitney U test was used for the two conditions with practice test findings, with the z-scores reported for this outcome). Repeated testing was taken into account by applying a Bonferroni correction for these tests. That is, with three tests on effects of condition conditions and six tests for comparisons across time, alpha was set at 0.005 (two-tailed). For effect size, we report the r-statistic (Field, 2005). This statistic tends to be qualified as small, medium, and large for respectively the values $r = 0.10$, $r = 0.30$, and $r = 0.50$.

6. Results

Prior experience and motivation. Table 1 shows the outcomes for the IEMQ. The students' overall mean scores are considerably below

Table 1
Mean scores[a] and standard deviations for task experience and motivation before training.

Condition	Task experience		Task relevance		Self-efficacy	
	M	*SD*	M	*SD*	M	*SD*
Practice-video ($n = 27$)	2.33	*0.96*	2.80	*1.57*	4.61	*1.25*
Video-practice ($n = 28$)	2.68	*1.29*	2.58	*1.36*	3.77	*1.56*
Video ($n = 27$)	2.09	*1.17*	2.18	*1.46*	3.63	*1.61*
Total (n = 82)	2.37	*1.16*	2.52	*1.47*	4.00	*1.53*

[a] Scale values range from 1 to 7; the scale midpoint is 4; a higher score indicates a higher appraisal.

the scale midpoint of 4 for both task experience and task relevance. The latter finding supports the design choice to include motivating before-after displays in the instruction booklet. There was no difference between conditions for task experience ($H = 4.39$, $p = 0.111$) and task relevance ($H = 4.51$, $p = 0.105$).

The overall mean score for self-efficacy is at the scale midpoint, indicating that the students began training with a moderate level of confidence in their capacity to accomplish the to-be-trained tasks. There was no statistically significant difference between conditions for self-efficacy, $H = 6.92$, $p = 0.043$.

Test scores before, during, and after training. Table 2 presents the findings for the various points at which the students were tested. The total mean score on the *pre-test* indicates that the students' prior knowledge regarding task performance was low (14.9% correct). There was no difference between conditions for the pre-test score, $H = 1.45$, $p = 0.485$.

The total mean score for *practice* was 66.1% correct. Compared to the pre-test this was a substantial improvement. The Wilcoxon signed rank finding revealed that the change was statistically significant, $z = 6.23$, $p < 0.001$, $r = 0.59$. This signals a large effect. Condition had a statistically significant effect on the practice scores, $z = 5.23$, $p < 0.001$, $r = 0.71$. This signals a large effect. The practice score for the video-practice group ($Mdn = 38.93$) was greater than for the practice-video group ($Mdn = 16.67$).

The finding that the practice-video group did as well as they did on practice was surprising, in view of the fact that these participants had not yet received any instructions for these tasks. This prompted an exploration of the data to find out whether there was any evidence of a carry-over effect from demonstrations viewed earlier in the tutorial. That is, the participants in this condition might have exploited some knowledge acquired from earlier task demonstrations and/or explorations to try out whether the same or similar objects could be used for the practice task (compare Stark et al., 2000). An illustration of such a carry-over effect comes from the data for indenting a citation on the left, the first task in the second chapter of the tutorial. The pre-test score on this time for the participants in the practice-video condition was 26%. During training, without having yet seen this task completed in the video, these participants achieved an 85% score on this task. What presumably contributed to this success are the earlier efforts to manipulate the “Left Margin” object on the ruler. Previous research had revealed that many participants struggled with the task of formatting the left margin of a document. The difficulty arises from the presence of four closely situated objects (i.e., “First Line Indent”, “Left Margin”, “Hanging Indent”, and “Left Indent”). Only a very delicate movement of the mouse yields the sought-after “Left Margin”. During task execution, participants therefore probably encountered the names and effects of one or more of the other objects. We assume that the participants in the present study struggled with formatting the left margin which was beneficial for formatting the left indent of a citation.

The total mean score for the *immediate post-test* was 61.3% correct. Compared to the pre-test this was a statistically significant improvement, $z = 7.75$, $p < 0.001$, $r = 0.61$. This signals a large effect. There was a decline from practice to immediate post-test that was not statistically significant, $z = -1.45$, $p = 0.147$. Condition had no effect on the immediate post-test, $H = 1.56$, $p = 0.459$.

The total mean score for the *delayed post-test* was 65.7% correct. Compared to the pre-test this was a statistically significant improvement, $z = 7.80$, $p < 0.001$, $r = 0.86$. This signals a large effect. There was an increase from immediate to delayed post-test that was not statistically significant, $z = 2.017.75$, $p = 0.045$. Condition had no effect on the delayed post-test, $H = 1.49$, $p = 0.474$.

Table 2
Mean percentage correct scores and standard deviations on pre-test, practice, immediate post-test, and delayed post-test.

Condition	Pre-test		Practice		Immediate post-test		Delayed post-test	
	M	*SD*	M	*SD*	M	*SD*	M	*SD*
Practice-video ($n = 27$)	14.4	*17.2*	45.4	*24.8*	56.5	*26.7*	60.2	*30.2*
Video-practice ($n = 28$)	13.0	*17.8*	86.2	*19.9*	64.7	*27.4*	66.5	*28.5*
Video ($n = 27$)	17.6	*18.8*	–	–	62.5	*29.2*	70.4	*26.7*
Total (n = 82)	14.9	*17.8*	66.1	*30.3*	61.3	*27.7*	65.7	*28.4*

7. Discussion

All in all, the results did not confirm our hypothesis for the first research question. We found no effect of *practice* on immediate or delayed recall. For both trained and transfer items, students in the video condition did as well as those who engaged in practice during training. The finding aligns with the outcomes from Hannafin and Colamaio (1987) and Baggett (1988), who studied video on assembly tasks and also found no effect of practice on learning.

The results also did not confirm our hypothesis for the second research question. We found no *timing* effect of practice on the outcomes on the two post-tests. This finding is surprising in that it does not support the prevalent view in education that students benefit most from a sequence where instruction precedes practice (e.g., Leppink et al., 2014; Reisslein et al., 2006). One explanation is that the students in the practice-video condition actually were motivated and well-prepared to study the video for what it might reveal about the solution method for the task they had just engaged in. The finding of a carry-over effect supports this contention.

The videos did well on the purpose for which software videos are generally designed, namely, to enable task completion. During training, students who first saw the video and then attempted to accomplish a task similar to what was modelled achieved an 86% success rate. This finding is similar to outcomes obtained in earlier studies involving these videos (van der Meij, 2014, 2017; van der Meij & van der Meij, 2016b). Because access to the video was prohibited during practice, the performance score for practice is primarily a sign of what participants learned from the video. Presumably with access to the video, there would be an even higher rate of successful, supported task completion.

That the practice scores in the video-practice condition were higher than in the practice-video condition is logical, because students in the latter group had yet to receive instructions for the practice task. Interestingly, the participants in the practice-video condition did significantly improve their task performance on practice tasks. This outcome was explained by a carry-over effect from having already seen a demonstration of performance on a related task.

The videos also did reasonably well as support for immediate and delayed recall of procedures. The students achieved significantly higher scores on both post-tests than before training. The pre-test mean score of 15% of tasks successfully completed rose to over 60% after training. The mean scores on the immediate post-test (ranging between 56.5% and 64.7%) meant a significant advance over the students' starting level, but they also left room for improvement.

One possibility for improvement could be the inclusion of instructions on selection rules. According to the Goals, Operators, Methods and Selection rules-model (GOMS) (Card, Moran, & Newell, 1983), it is important for users not only to learn a procedure or method, but also to get to know when a method can or cannot be applied, and which method is preferable when different methods can achieve the same end result. This choice of methods is captured in the selection rules. Because selection rules were not taught in the video instructions, and the tests presented students with all trained tasks at once, testing may have presented the students with a situation challenging them not only to reproduce the correct procedural steps, but also primarily to select the proper method for each task.

Another possibility for improvement could come from inclusion of varied practice tasks that further assist students in learning to select the proper solution method. However, this option holds the risk that performance during training could become less successful. That is, the study by Helsdingen, van Gog, and van Merriënboer (2011) showed that varied practice tasks yielded greater learning after training, but this came at the cost of diminished success on the tasks practiced during training. A similar finding for practice organization has also been reported for procedural knowledge development (Schmidt & Bjork, 1992), and for motor skills development (Bangert, Wiedemann, & Jabusch, 2014). An important reason for being concerned with lower success of practice tasks is the possibility of attrition. It is conceivable that there is a greater risk of early drop-out from a video tutorial when participants experience failure in achieving modelled tasks.

In addition to enhancing performance on trained tasks, instructional videos should also facilitate transfer. Videos should enable participants to solve new but related problems after training. Unfortunately, the transfers test had poor reliability and therefore no transfer findings could be reported.

8. Conclusion

Software demonstrations generally must serve the dual goals of supporting both doing and learning. This is a challenging design task, because there is an inherent tension between providing support that facilitates task completion and stimulating reflection to enhance learning. Inadequate support for doing may result in loss of user interest in the instructions. Too little support for learning may result in continued dependency on documentation. In short, it is important to distinguish between temporary or transitory and permanent effects of training. Instructional measures that maximize performance during training can be sub-optimal for learning and transfer. The present study shows that videos for software training can be designed to achieve both goals with good to moderate levels of success.

In educational settings, software training can and should be optimized for learning without sacrificing too much to achieving performance success during training. The DBT-model for software training recently presented by Brar and van der Meij (2017) provides a large number of instructional features for doing so. One of these measures, the inclusion of practice, was tested in the present study. No evidence was found that learning from software training was enhanced by this inclusion. What caused the absence of the anticipated contribution of practice? Why were the practice-video, video-practice, and control conditions all equally effective? Finding the answers to these questions is important, so that we can further improve the design of videos for software training.

Compliance with ethical standards

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

Funding

No funding was obtained for this study.

Conflicts of interest

The authors declare that they have no conflict of interest.

References

Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research, 70*(2), 181–214. https://doi.org/10.3102/00346543070002181.

Baggett, P. (1988). The role of practice in videodisc-based procedural instructions. *Transactions on Systems, Man and Cybernetics, 18*(4), 487–496. https://doi.org/10.1109/21.17366.

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

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

Bangert, M., Wiedemann, A., & Jabusch, H. C. (2014). Effects of variability of practice in music: A pilot study on fast goal-directed movement in pianists. *Frontiers in Human Neuroscience, 8*, 1–22. https://doi.org/10.3389/fin-hum.2014-00598.

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*. London, UK: Lawrence Erlbaum Associates.

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.

Eccles, J. S., & Wigfield, A. (2002). Motivation beliefs, values, and goals. *Annual Review of Psychology, 53*, 109–132. https://doi.org/10.1146/annurev.psych.53.100901.135153.

Ertelt, A. (2007). *On-screen videos as an effective learning tool. The effect of instructional design variants and practice on learning achievements, retention, transfer, and motivation* (Unpublished doctoral dissertation). Germany: Albert-Ludwigs Universität Freiburg.

Field, A. (2005). *Discovering statistics using IBM SPSS statistics* (2 ed.). London, UK: Sage.

van Gog, T. (2011). Effects of identical example-problem and problem-example pairs on learning. *Computers & Education, 57*, 1775–1779. https://doi.org/10.1016/j.compedu.2011.03.019.

van Gog, T., Kester, L., & Paas, F. (2011). Effects of worked examples, example-problem, and problem-example pairs on novices' learning. *Contemporary Educational Psychology, 36*, 212–218. https://doi.org/10.1016/j.cedpsych.2010.10.004.

van Gog, T., & Rummel, N. (2010). Example-based learning: Integrating cognitive and social-cognitive research perspectives. *Educational Psychology Review, 22*, 155–174.

Grabler, F., Agrawala, M., Li, W., Dontcheva, M., & Igarashi, T. (2009). Generating photo manipulation tutorials by demonstration. *ACM Transactions on Graphics, 28*(3), 66.61–66.69. https://doi.org/10.1145/1531326.1531372.

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.

Hannafin, M. J., & Colamaio, M. A. E. (1987). The effects of variations in lesson control and practice on learning from interactive video. *Educational Technology Research and Development, 35*(4), 203–212. https://doi.org/10.1007/BF02766965.

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. https://doi.org/10.1037/a0022370.

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.

Kerr, M. P., & Payne, S. J. (1994). Learning to use a spreadsheet by doing and by watching. *Interacting with Computers, 6*(1), 3–22.

Leppink, J., Paas, F., van Gog, T., van der Vleuten, C. P. M., & van Merriënboer, J. J. G. (2014). Effects of pairs of problems and examples on task performance and different types of cognitive load. *Learning and Instruction, 30*, 32–42. https://doi.org/10.1016/j.learninstruc.2013.12.001.

Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology, 84*(4), 429–434.

Paas, F., & Van Merriënboer, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Educational Psychology, 86*(1), 122–133.

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

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

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.

Salden, R., Koedinger, K. R., Renkl, A., Aleven, V., & McLaren, B. M. (2010). Accounting for beneficial effects of worked examples in tutored problem solving. *Educational Psychology Review, 22*, 379–392. https://doi.org/10.1007/s10648-010-9143-6.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*, 207–217.

Stark, R., Gruber, H., Renkl, A., & Mandl, H. (2000). Instruktionale Effekte einer kombinierten Lernmethode: Zahlt sich die Kombination von Lösungsbeispielen und Problemlöseaufgaben aus? [Instructional effects of a combined learning method: Does the combination of worked-out examples and problem-solving tasks pay off?] *Zeitschrift für Pädagogische Psychologie, 14*, 205–217. https://doi.org/10.1024//1010-0652.14.4.206.

Sweller, J., & Cooper, G. A. (1985). The use of worked examples as s substitute for problem solving in learning algebra. *Cognition and Instruction, 2*(1), 59–89. https://doi.org/10.1207/s1532690xci0201_3.

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

van der Meij, H. (2007). Goal-orientation, goal-setting and goal-driven behavior in (minimalist) user instructions. *IEEE Transactions on Professional Communication, 50*(4), 295–305.

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

van der Meij, H. (2017). Reviews in instructional video. *Computers & Education, 114*, 164–174. https://doi.org/10.1016/j.compedu.2017.07.002.

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., Karreman, J., & Steehouder, M. (2009). Three decades of research and professional practice on software tutorials for novices. *Technical Communication, 56*(3), 265–292.

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. (2016). 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.

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