The effect of timing of practice in video-based software training

Tessa Voerman

Educational Science and Technology

University of Twente

First supervisor: Hans van der Meij

Second supervisor: Alieke van Dijk

February 2020

Table of contents

Abstract

This study examined the effectiveness of timing of practice on learning from video instruction for software training ('videos'). An experiment is reported in which 53 university students received an alternation of videos complemented with practice as an instructional feature ('training'). The videos were created using guidelines from multimedia and Demonstration-Based Training research. The following variables were measured: engagement, task relevance, self-efficacy and conceptual and procedural knowledge. The participants were randomly distributed among three conditions. One condition performed a practice activity before watching the tutorial (PT), another the other way around (TP), and the third practiced at both moments in time (PTP). As PTP combines the strengths of both PT and TP, this condition was expected to outperform the other conditions. Video logging showed that almost all videos were watched. Questionnaires before and after training and a knowledge test after training showed an enhancement of task relevance, self-efficacy and conceptual and procedural knowledge scores, attributed to the notable quality of the videos. The timing of practice significantly increased self-efficacy change and procedural knowledge scores. An explanation for this is the availability of external feedback, which enabled the learners' evaluation and improvement of their performance and therefore learning. Other significant findings are the pairwise comparisons of the TP- and PT-conditions on both self-efficacy growth and procedural knowledge scores. PTP was expected to outperform PT and TP, but barely did so. Future research is needed to create a better understanding of under which circumstances timing of practice is most effective in software training.

Keywords: *video, demonstration-based training, practice, software*

1. Introduction

Modern-day organizations increasingly need task performance trainings that are effective on learning and readily accessible. One way to facilitate this need are video-based instructions that model task accomplishment. A prominent approach for such video creation is Demonstration-Based Training (DBT), in which the learner observes a task or procedure that is demonstrated and explained by a model in order to understand the procedure and to reproduce this in the future (van der Meij & van der Meij, 2014). This "acquiring of new patterns of behavior through observing the behavior of others" (Bandura, 1977, p.3) is hereinafter called 'learning'. DBT's are typically video-centered, as videos suit organizations' need for effective and readily accessible training. This usefulness increases videos' popularity in educational settings (Kay & Kletskin, 2012), leading to a growing body of literature that recognizes the importance of thoroughly designed DBT for enhancing learning. Specifically, the research of van der Meij (2018) on DBT based videos will be relevant to the current study, as will be discussed later.

DBT based videos generally aim to achieve the dual purpose of supporting task achievement and learning. The latter enables the learner to perform tasks independently (Grabler, Agrawala, Li, Dontcheva & Igrashi, 2009; Kay & Kletskin, 2012). To maximize this learning effect, it is necessary to add instructional features to the design of the demonstration (Rosen, Salas, Pavlas, Jensen, Fu & Lampton, 2010; van Gog, 2011). These can be activities and information, such as the use of voice-overs, examples or practice moments. These instructional features influence observational learning processes such as *attention*, *retention*, *production* and *motivation* (see Bandura, 1977). Research that attempts to provide insight into the variety of instructional features and their effects on observational learning processes is scarce. Therefore, more systematic research is necessary to understand in what circumstances which instructional features are most effective for enhancing learning (Rosen et al., 2010).

One particular instructional feature that is scarcely investigated, but has the potential for benefitting learning, is *practice*. Practice is an activity in which the learner performs procedural tasks for task achievement and learning purposes. In the context of Bandura's observational learning processes, practice enhances production, leading to a behavioral outcome in which the learner shows to have learned the observed behavior by producing it independently from the model (Bandura, 1977; Brar & van der Meij, 2017).

According to the few empirical studies in the field of software training available, the influence of practice on learning is sensitive to the moment that it occurs vis-a-vis the demonstration (van der Meij, Rensink & van der Meij, 2017; van der Meij, 2018). This timing of practice plays a pivotal role in the current study. That is, the current study is set out to compare three scenarios of timing of practice, each having their own perks towards learning. These scenarios have been proposed by the earlier study of van der Meij (2018) on DBT videos in software training. The following paragraphs will elaborate on the effect of timing of practice on learning by portraying the three scenarios.

1.1 Scenario 1: Tutorial-Practice

In the most common and frequently occurring scenario, practice directly follows a tutorial (van der Meij & Dunkel, 2020), which we call the tutorial-practice condition (TP). When looking at the DBT-instructional-features categorization of Rosen et al. (2010), the practice component in TP can be categorized as a retrospective activity, in which the learner cognitively reviews the just-observed task.

TP has two advantages. First, retention is stimulated by simply processing the task once again (Rosen et al., 2010; Sweller & Cooper, 1985), this time in a cognitively more active manner compared to observation. Second, motivation, retention and production are stimulated as the learner checks whether he has understood the task. In this process, the learner draws his attention to those aspects in his cognitive model of the task where mistakes may have been made and need further revision (van der Meij & Dunkel, 2020). Such reflective self-regulation during practice is likely to indicate the learner's awareness of its current and desired capabilities towards the task. This awareness can motivate re-examination of the task (van der Meij, van der Meij, Voerman & Duipmans, 2017; van Gog, 2011), which is likely to result in a deeper understanding of the task.

1.2 Scenario 2: Practice-Tutorial

A second less-researched scenario places the timing of practice before the tutorial. An example of this is the realistic situation in which a software user runs into a problem and turns to a tutorial for an explanation and solution. This practice-tutorial condition (PT) is characterized by the unguided exploration of the software environment as a starting point, followed by instruction.

PT has three advantages. First, practice before a tutorial is a preparatory activity (Rosen et al., 2010) that requires the learner to think hard about the problem. This is likely to generate a sense of connection to the problem and activate prior knowledge, which allows for connection making between new and existing information during the tutorial observation (Kalyuga, 2017, in van der

Meij, Rensink & van der Meij (2017). Furthermore, prior knowledge facilitates the user in interpreting and directing their experiences (van der Meij & Gellevij, 2004). Such processing is likely to increase retention and production (Stark et al., 2000, in van Gog, 2011).

Second, practice can make the learner aware of a lack of knowledge and skills needed in the task execution. This awareness arises from external feedback, as the software interface tells the user which actions lead towards and away from the goal. This is somewhat similar to the reflectivity when practicing after the tutorial, but in this case, it may not nearly be as clear to the learner what these shortcomings are. After all, the learner is not yet familiar with the correct procedure. Such awareness is likely to result in a desire for task understanding. From this point of view, practice before the tutorial can stimulate internal feedback. Because of this desire, the learner can become stimulated to observe the instruction actively and in a concentrated state (Stark et al., 2000 in van Gog, 2011; van Gog, Kester & Paas, 2011).

Third, practice before the tutorial can enhance perceived task relevance. That is, when a learner is able to place the observed procedure (encountered during the tutorial) in the just-experienced context (encountered during the practice before), this can have a positive effect on motivation and therefore learning (van der Meij & van der Meij, 2013). However, when this placement fails due to weak relevance, motivation and learning can be hindered. An example of this is when the learner unknowingly has practiced in an unrelated part of the software environment and therefore has trouble linking the procedure to the experienced context.

1.3 Scenario 3: Practice-Tutorial-Practice

In the PT scenario, a software user ran into a problem and tried to find a solution by watching a tutorial. In the most realistic sense, this user does not stop there, but then tries to apply the new knowledge in the software environment. This leads to a third, and rarely investigated, scenario: the practice-tutorial-practice condition (PTP), where practice activities are involved both before and after the tutorial.

Besides this attempt to recreate a realistic usage of software environments in an experimental setting, the idea behind using PTP as an experimental condition is that PTP combines the strengths of both TP and PT. This leads to two advantages. First, PTP is likely to activate prior knowledge (PT). Second, it can be expected that PTP stimulates active task processing during both practice (TP) and tutorial (PT) activities, as both activities prompt the learner’s reflection on the task.

This way, a learner can be motivated twice to deepen her or his understanding of the task. Because of these combined advantages, one might expect PTP to be more effective than TP and PT.

PTP as an experimental condition has been pioneered by van der Meij (2018), with similar expectations of PTP being advantageous towards PT and TP. However, the potential of PTP has been contradicted in the results in the research of van der Meij (2018), as the PTP-condition had lower test scores than the TP- and PT-condition on an immediate knowledge test. It has been speculated that the found result is due to the absence of external feedback (Fiorella & Mayer, 2018). In software training, the user can learn from the internal feedback that the software gives to the user's actions. For example, when clicking a button, a window pops up, but it is not the window the user wanted to find. Fiorella and Mayer (2018) suggest that this internal feedback is not sufficient and that users need additional, external feedback to be able to evaluate and improve their performance. To facilitate learners with such evaluation using feedback, the current study includes the possibility for video replay in the second moment of practice in the PTP-condition. This inclusion is an important modification on the research approach of van der Meij (2018).

2. Research Design and Research Questions

The current study has an experimental design and investigates the influence of three scenarios of timing of practice (the three conditions TP, PT, PTP) on a set of dependent variables that will be explained further on. The following research question is central:

'To what extent does timing of practice influence the users' viewing actions, motivation and learning from video-based tutorials for software training?'

More specifically, the subordinate research questions are as follows:

1. What is the role of timing of practice on viewing, henceforth called engagement, of the videos?
2. What is the role of timing of practice on motivation?
3. What is the role of timing of practice on learning?

The first subordinate research question will be answered by examining the dependent variable *engagement*. Engagement is measured to study how much time a participant spends watching the video tutorials. Engagement is an indicator of quality of the videos that are used in the study. It is important to have a grasp of the study's video quality, as having poor quality videos may blur study results. To ensure the quality of the videos, their design is incorporated with instructional features, as will be explained in paragraph '3.2: Materials: Instructional videos'.

Returning to engagement in a more specific manner, engagement indicates the number of seconds of a total video that are watched completely, skipped or replayed. As a result, it can be marked which parts of the videos were insightful or problematic to the learner. Engagement is measured in two ways. First, *unique play* tells the number of seconds the video has been set in play-mode and thus is likely to have been watched. Second, *replay* tells the proportion of the video that has been replayed.

The second subordinate research question on motivation will be answered by examining the two key elements of the Expectancy-Value Theory (Wigfield & Eccles, 2000) namely *task relevance* and s*elf-efficacy*.

Task relevance is the perceived importance of a task. The higher the score on task relevance, the more valuable the task is to the learner, resulting in a stronger motivation to more frequently complete the task (Bandura, 1977; Wigfield & Eccles, 2000). Earlier studies on practice in software education (e.g. van der Meij, 2018) have not found significant changes in task relevance. However, it is expected that the current study is likely to increase task relevance through optimal video design and the use of an adult population, for example compared to the children used in van der Meij (2018). Perhaps these adults use software environments like the one used in the current study more regularly and are therefore more sensitive to task relevance changes. Therefore, it might be interesting to examine task relevance.

Self-efficacy is the learner's perceived confidence in their capacities in relation to the task. The higher a self-efficacy score, the more motivated the learner is likely to be to execute the task and use it in practice (Bandura, 1977; Wigfield & Eccles, 2000). As practice is likely to positively influence self-efficacy, it is expected that the double practice activities of the PTP-condition will score higher on self-efficacy than the PT and TP conditions.

The third subordinate research question regards the timing of practice on learning as a dependent variable. Learning is a broad term that can be distinguished in *conceptual knowledge*, that shows how much conceptual information is remembered, and *procedural knowledge* that shows how well the learner can execute the just-learned task without help (Brar & van der Meij, 2017).

The measurement of conceptual knowledge has not been executed in other studies regarding the timing of practice (e.g. van der Meij et al. 2017; van der Meij 2018). The inclusion of conceptual knowledge in the current study is likely to contribute to the body of knowledge as to how practice can function optimally as an instructional feature in software training.

The measurement of procedural knowledge has been executed in the beforementioned practice studies, but with contradicting results. As more empirical research is needed to gain more consensus on the potential of practice as an instructional feature, the current study examines procedural knowledge as well. Following the line of thought that as practice can be positively influential on procedural knowledge scores, it is expected that the double practice activities of PTP will result in higher procedural knowledge scores than among the PT and TP conditions, those having only one practice activity.

3. Method

The experiment consisted of an alternating sequence of practice and video watching activities. This ‘training’ was followed by a knowledge test. Questionnaires were used throughout the duration of the experiment. This chapter elaborates on the specifics of the experiment.

3.1 Participants

The 53 participants of this study were students from a Dutch university. Their mean age was 24.28 years old (*SD* = 2.53). Rather than children that participated in the earlier study of van der Meij (2018), this experiment involved adults that were deemed to have more than sufficient self-regulatory skills for the training, which can enhance learning (Bandura, 1977). The participants were randomly distributed over the three conditions. All participants had experience in working in the software environment and they were proficient in English, allowing for all instructional materials to be in English.

3.2 Materials

3.2.1 Instructional videos

The materials used during training consisted of seven videos on Word. They were especially created for this study and made accessible through a website. The first video was an introduction, which roughly explained the learner what to expect in the other six videos and perhaps more importantly, the extent to which the learner could use the tasks in daily life. Its goal was therefore to increase motivation and grab the learner’s attention.

The other six videos were tutorials that demonstrated a procedural stepwise progression until task completion, hereinafter called ‘videos’. The videos were grouped in three chapters. The first chapter presented two tasks on typing less in Word while still getting the same result (Using acronyms to shorten typing core phrases (total duration: 136 seconds); Using AutoText to have Word suggest phrases (112 seconds)). The second chapter showed two tasks on automatically numbering figures and tables (Creating Captions to ensure accuracy of Figures (or Tables) (139 seconds); Cross-referencing to a Figure (91 seconds)). The third chapter presented two tasks on how Word can be used to stick information together that should not be separated by line or page ends (Creating nonbreaking spaces for statistical outcomes (88 seconds); Keeping together header and paragraph (71 seconds).

The main structure of all the videos was the same. After an introduction part, the demonstration of the task procedure started. These procedures were demonstrated by a model that executed the task flawlessly (van der Meij & van der Meij, 2013). This demonstration was a visual task completion in the software environment, supported by a narrative, as "people learn more deeply when they build connections between a verbal representation and a pictorial representation of the same material" (Mayer, 2008, p.755). The narrative contained conceptual information and action steps that told the user what to do, why and possibly where to find needed material, presented by a verb followed by a noun (e.g. "Click Okay to go back to your document") (van der Meij & Gellevij, 2004; van der Meij et al., 2017). These action steps were executed by the model in a just-in-time matter. The videos ended by pointing out the end result that refers to the title and the goal of the video (e.g. "Notice that the heading and paragraph now stick together").

The design of the videos was based on the cognitive theory of multimedia learning (Mayer, Heiser & Lonn, 2001). This widely accepted theory illustrates how multimedia learning operates by opting three main assumptions: dual channeling, limited capacity and generative learning (Mayer et al., 2001). The *dual channel assumption* entails that visual and verbal information are processed separately via different sensory channels in the brain. When visual and verbal information enter the learner's sensory channels, these systems select which information is worth processing further into a visual and verbal model wherein the information is organized (Bandura, 1977). The reason for this selection is the *limited capacity assumption*, which implies that these channels are only able to process a small amount of verbal or visual information at once. Then, prior knowledge is activated, resulting in an integration of new information and already existing information. The result of this integration is called knowledge, which is now ready to be stored in long-term memory. Overall, the level to which the learner can cognitively process such information into knowledge is a key factor in its ability to learn deeply, which is called the *generative learning assumption* (Mayer et al., 2001).

However, to be able to learn from multimedia, the learner needs to focus its *attention* to the video in order to recognize essential information from the model's demonstration (Bandura, 1977). A more obvious requirement for observational learning is that the learner needs to remember the demonstration, there must be *retention* (Bandura, 1977). Furthermore, the learner needs to be able to exhibit the observed behavior; the student should show that *(re)production* is obtained (Bandura, 1977). Even more, the learner is only able to learn from video if it entails the *motivation* to do so (Bandura, 1977).

These four learning processes, attention, retention, production and motivation, are key in observational learning, as described in Bandura's Social Cognitive Theory (1977). To maximize learning from videos in the current study, instructional features were used in the design of the videos. For an overview on which instructional features can influence which observational learning process see Figure 1. The following paragraphs will elaborate on which instructional features were used in the video design of the current study, categorized per learning process.

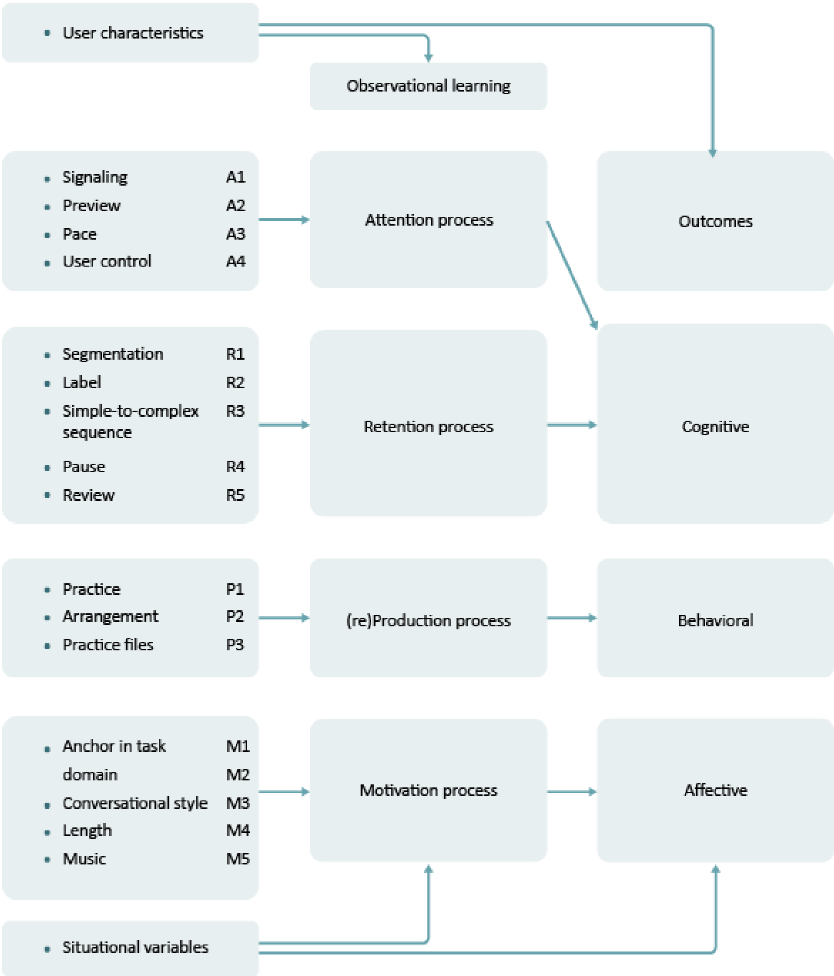


Figure 1. *DBT-model of the connection between conditions, instructional features, learning processes and outcomes in software training* (Brar & van der Meij, 2017).

Attention. Over the past decades, research has established several guidelines for getting and keeping the learner attentive to important information to complete a task. By providing the learner with guidance on where to focus, they are supported in making the distinction between important and less important information (Grossman, Salas, Pavlas & Rosen, 2013). One instructional feature that impacts attention is the *pace* of the videos (A3). That pace should be suited to the cognitive level of the participants to keep cognitive load appropriate for the target group (van der Meij & van der Meij, 2016; Koumi, 2003). That is, if the pace is too slow, the learner is likely to get bored or inattentive, whereas a too fast pace would overwhelm them (van der Meij & van der Meij, 2013). Another attention-supportive feature is visual cues such as zooming and *signaling* (A1). These techniques were used to steer attention to relevant information and refrain the learner from being able to get distracted by irrelevant information (Mayer, 2008; Grossman et al., 2013). An example of signaling is circling of buttons in the software environment in a red color, which is a commonly used and popping color for drawing attention (Kosslyn, Kievit, Russell & Shephard, 2012). A third feature that enhances attention is the use of *previews* (A2) at the start of each video. The previews pointed out the initial problem state and the goal of the video. By preparing the learner for what is to be expected in the video, prior knowledge can be activated and attention to the procedure is reinforced (Fiorella & Mayer, 2019). A fourth attentional feature is user control (A4). The learner can influence the pace of the video by starting, pausing and rewinding the video with a toolbar.

Retention. Several guidelines have been used in the video design to enhance the learner's storage of information into long-term memory for future use. For example, the videos were titled, (R2), revealing the intended goal and stimulating prior knowledge (Fiorella & Mayer, 2019). Another example is the extensive use of segmentation (R1), which involves dividing a long fragment of video into more manageable chunks (Mayer, 2008; Fiorella & Mayer, 2018). Segmentation strengthens the learner's cognitive model of the task (van der Meij & van der Meij, 2013). For example, each video was divided into sub goals, which is recommended when a learner must deal with many actions (van der Meij & Gellevij, 2004). Each sub goal was introduced by on-screen text, stating its title (R1, R2), which lasted four seconds and did not contain audio (see Figure 2). All goals and sub goals reflect the task structure and are presented in gerund-form (van der Meij & Gellevij, 2004).

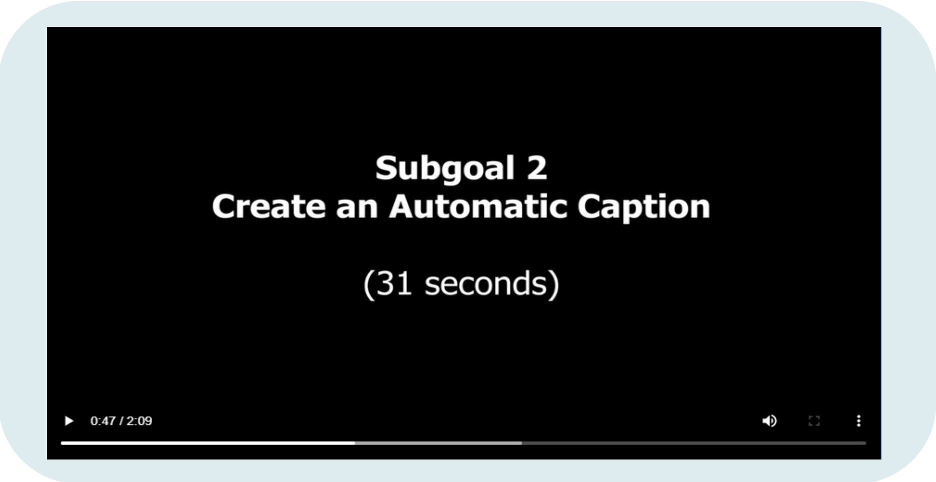


Figure 2. *Screenshot from video showing the title of the sub goal and the duration of the upcoming part.*

The segmentation frees up cognitive capacity for processing the next part of the tutorial (Mayer, 2008; van der Meij, 2018). Also, segmentation functions as a natural pause (R4) in the instruction, each lasting two seconds, which allow the learner to take in the observed procedure

(Spanjers, van Gog, Wouters, & van Merriënboer, 2012; van der Meij & van der Meij, 2013; Fiorella & Mayer, 2018). These small moments of rest can encourage self-reflection on whether the information has been understood and whether revision is necessary (Koumi, 2003). This way, cognitive load is kept manageable, supporting the transition from working memory to long-term memory storage (Sweller et al., 1998; Mayer, 2008; Koumi, 2003).

Production. Learning from an observed demonstration can be exhibited by (re)producing the task achievement (Brar & van der Meij, 2017). As can be seen in Figure 1, production solely derives on *practice* (P1), *arrangement of practice* (P2) (in this study called 'timing of practice'; see Chapter 1: Introduction) and practice-facilitating *practice files* (P3), which were Word-environments with marked problems that the participant needed to solve by executing the task. The practice files were structured similarly to how the Word environment looks in the videos, but they differed in superficial characteristics to have the learner take a more active approach in understanding the task (compare van Gog, 2011). To maximize understanding, the practice files kept irrelevant and distracting information to a minimum.

Motivation. Learning only occurs when it is valuable and worthy to the learner. Motivation is underlying to the processes of attention, retention and production (van der Meij & van der Meij, 2016). The design of the videos supports motivation by *anchoring the tasks in the task domain* (M1), which enlarges perceived value of those tasks (van der Meij et al., 2017). That is, each video demonstrated a task on how Microsoft Word can be applied for efficient writing of theses and explicitly mentioned so in its narrative. This way, the task in the video is presented in its actual interface, which is likely to enhance learning (van der Meij & van der Meij, 2013). Furthermore, the narrative of the videos used *conversational style* (M2) by using words such as 'you' that enhance personification (Clark & Mayer, 2003; van der Meij, 2018). This conversational style enhances learning more than a formal style would, as the learner get more connected to the narration and have to put in less effort to create an understanding of the narration (Mayer, 2008). Besides that, each video *length* was relatively short (range: 71-136 seconds) (M3), to keep the learner motivated and prevent attention from slacking and cognitive overflow (Guo, Kim & Rubin, 2014).

3.2.2 Instruction booklets

Paper-based instruction booklets navigated the participants through the sequence of tutorial watching and practice activities. The booklet, based on the ones used in van der Meij (2018), instructed to engage in a practice activity for a minimum of 30 and a maximum of 60 seconds and to move on when stuck (see Figure 3). This minimum was to ensure that practice activities occurred.

The maximum was meant to keep up the pace of the experiment, as the practice tasks were short and could be completed within the set time frame. The decision to include a maximum was made to try to prevent participants from becoming unmotivated due to extensive runtime of the experiment or constant failure in the practice activity. The actual duration of practice was not measured due to practicality issues.

Step 3.1.2 **Watch** the video **'3.1 Creating nonbreaking spaces for statistical outcomes'**. *You can watch the video as many times as you like.* When you are done, move on to step 3.1.3.

Step 3.1.3. On page 5, **create nonbreaking spaces** *for a statistical outcome.*

Do NOT perform any activities regarding previously learned information, for the sake of this experiment.
You are allowed to return to video 3.1.
Practice for at least 30 seconds, but no longer than a minute.

Figure 3. *Excerpt from instruction booklet for TP-condition*.

Also, as seen in Figure 3, in the TP-condition, the booklet reminded the participant that it was allowed to replay the just-viewed tutorial during practice. This is a major difference from the research of van der Meij (2018), where replaying was prohibited and thus feedback was not available. The study of van der Meij (2018) received criticism on the absence of feedback (Fiorella & Mayer, 2018).

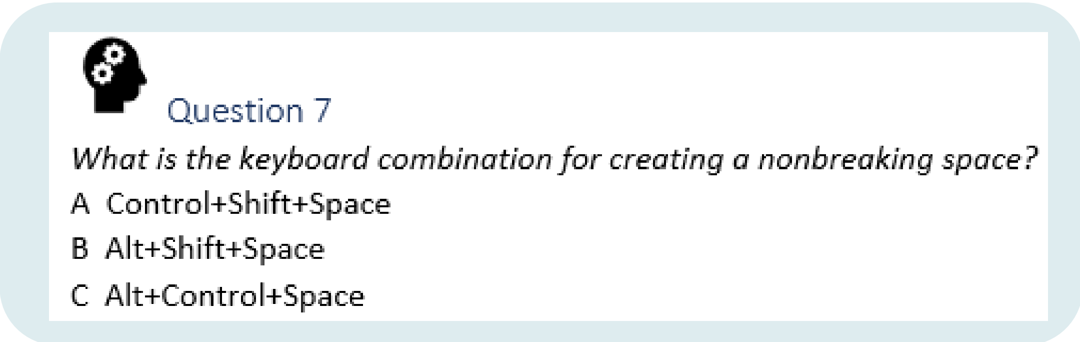
3.2.3 Instruments

The instruction booklets also contained questionnaires and the instrument for handling the knowledge test. The questionnaires are comparable to those used in van der Meij (2018), that in turn were based on the Fragebogen Actuelle Motivation (Rheinberg et al., 2001). This paragraph describes each questionnaire and the test instrument individually.

The *Initial Experience and Motivation Questionnaire* (*IEMQ*) asks three questions on prior experience, task relevance and self-efficacy. It was to be completed after reading a short description of a task and before engaging with each of the six tasks. The first question examined the prior experience of the learner: "Do you ever have to do this task?". The second question asked about the task relevance: "How often will you need to complete this task in the future?". Answers were given on a 5-point Likert scale. The items were reverse coded for reliability purposes (Hartley & Betts, 2013), being 1= very often and 5 = never. The third question measured self-efficacy to the task (Bandura, 2006): "How well do you think you can complete this task?". Again, a 5-point Likert scale was used. Cronbach's alpha reliability analyses indicated good reliability for initial task relevance (α = 0.744) and self-efficacy (α = 0.885).

The *Final Motivation Questionnaire* (*FMQ*) examines the learner's appraisal of task relevance and self-efficacy after training. There were six questions on task relevance, one for each practiced task (e.g. "I find it important to be able to cross-reference to a Figure"). Self-efficacy was appraised similarly (e.g. "I can now cross-reference to a Figure"). The items were reverse scored on a 5-point Likert scale, likewise to the IEMQ. Reliability was calculated using Cronbach's alpha, which showed satisfactory results for task relevance (α = 0.682) and self-efficacy (α = 0.762).

A *knowledge test* assessed conceptual and procedural knowledge for each of the six trained tasks. The six items on conceptual knowledge were multiple-choice questions, flagged by a picture of a thinking head (see Figure 4).


Question 7

What is the keyboard combination for creating a nonbreaking space?

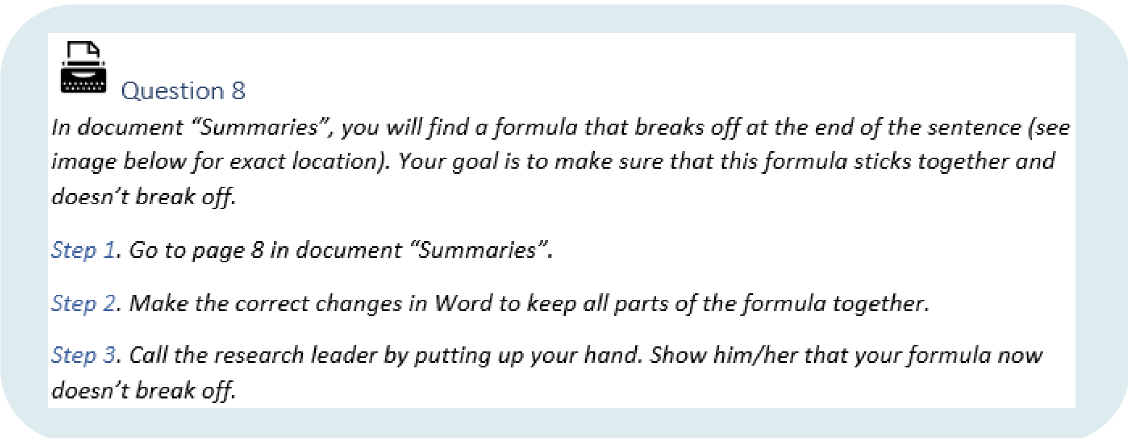
A Control+Shift+Space

B Alt+Shift+Space

C Alt+Control+Space

Figure 4. *An example of a conceptual knowledge question exerted from the instruction booklet*.

The six items on procedural knowledge were flagged with a typewriter, as they were hands-on tasks that had to be executed in the software environment (see Figure 5). For all test tasks, a Word file was prepared, which included the same underlying problems and structure as demonstrated in the tutorials, varying only in surface features.


Question 8

In document "Summaries", you will find a formula that breaks off at the end of the sentence (see image below for exact location). Your goal is to make sure that this formula sticks together and doesn't break off.

Step 1. Go to page 8 in document "Summaries".

Step 2. Make the correct changes in Word to keep all parts of the formula together.

Step 3. Call the research leader by putting up your hand. Show him/her that your formula now doesn't break off.

Figure 5. *An example of a procedural knowledge question exerted from the instruction booklet.*

Scoring of knowledge test items. Test performance was scored using a codebook. No points were rewarded for wrong multiple-choice answers and/or failing to execute the procedure correctly. A score of 1 was rewarded for a correct multiple-choice answer and/or being able to show correct procedure execution on the computer. For both conceptual and procedural learning, the maximum score was 6.

View logging. As a proxy for viewing, an unobtrusive logging program was used. The program computed the number of seconds the video was played at least once. It then was expressed as a percentage of the length of the video in seconds. So, 0% meant that no part of the video was ever set into play mode. When every unique second of a video had been set in to play mode at least once, a maximum score of 100% was reached. This engagement was computed for all videos together. Also, the program computed the number of seconds the video was replayed. As replay rarely occurred, this study will not further elaborate on this.

3.3 Procedure

The study was conducted in researcher-led sessions, in which one to four participants could participate at the same time. Each session took about an hour. At the start of the session, participants received information that they would engage in Word-software training, to assist them in writing their reports and theses more efficiently. Furthermore, they received instruction about the use of the instruction booklets and the practice files. The total introduction to the experiment lasted about 5 minutes. Next, the participants were instructed to start the training by following the instruction booklet precisely. The booklet asked the participants to fill in the IEMQ, watch the tutorials and engage in practice activities (P1) related to the tasks of the tutorials. In this booklet, the participants were instructed to view the tutorial as often as they liked, until they felt sure they could reproduce the task procedure. Also, during practice, the participants were allowed to replay the videos. The participants started by watching an introductory video, so they knew what to expect. The total training took about 40 minutes in the PT- and TP-conditions and about 50 in the PTP-conditions.

After a small break (1 minute), the participants moved on to the conceptual and procedural knowledge test. This test was also instruction-booklet led and took about 10 minutes. During this test, it was forbidden to consult the tutorials. After each procedural knowledge question, the participants were asked to call for the research leader, who immediately scored the correctness of the executed procedure. After completion of the knowledge test, the participants filled in the FMQ to close the experiment session.

3.4 Analysis

When checking for random distribution of participants over conditions, no significant differences were found for age (F(2,50) = 0.892, p = .416), the language of the participants' Word environment (F(2,50) = 0.704, p = .500) and frequency of prior task execution (F(2,50) = 2.049, p = .140). This latter means that all participants had similar prior knowledge of the tasks at hand. However, gender was not normally distributed per condition (F(2,50) = 4.378, p = .018). General Linear Model tests were used to check the potential influence of this covariate (see Chapter 4. Results). Both variables *unique play* and *replay* are not normally distributed and call for nonparametric tests such as Kruskall-Wallis. Initial checks for assumptions of variance homogeneity (Levene's tests) showed no significant differences between conditions regarding any variable. When checking for random distribution, there was no difference between conditions on initial *self-efficacy* (F(2,50) = 2.904, p = .064) or initial *task relevance* (F(2,50) = 0.347, p = .709).

ANOVAs were used to assess differences among conditions on motivation and learning. To compute effect sizes, Cohen's d (Cohen, 1988) was used. According to this measurement, an effect size is small when d = 0.2, medium when d = 0.5 and large when d = 0.8.

4. Results

4.1 Engagement

The findings on *engagement* are presented in Table 1. The videos have been watched almost completely (96.79%). A Kruskall-Wallis *H* test showed no statistically significant difference in *unique play* scores between the conditions (*H* = 5.500 and *p* = .064). As replay hardly occurred, further analyses did not take place.

Table 1

Means for Engagement[a] by Condition

	Engagement	
	Unique Play	
Condition	*M*	(*SD*)
TP (*n* =18)	96.34%	(6.05)
PT (*n* = 18)	98.50%	(3.90)
PTP (*n* = 17)	95.15%	(9.54)
Total (n = 53)	*96.79%*	*(6.88)*

[a] in percentages

4.2 Motivation: Task Relevance

A repeated measures ANCOVA was carried out with condition as the independent variable and the two *task relevance* scores as the dependent variables. Gender was a covariate but had no effect on the results. Table 2 shows that there was a slight, non-significant increase in task relevance in all conditions, corrected for gender, $F(2,50) = 2.68$. $p = .766$. However, there was no effect of condition found.

Table 2

Means for Task Relevance[a] by Condition

	Task Relevance					
	Before[a]		After[a]		Growth	
Condition	*M*	(*SD*)	*M*	(*SD*)	*M*	(*SD*)
TP (*n* =18)	3.34	(0.62)	3.87	(0.63)	0.60	(0.65)
PT (*n* = 18)	3.16	(0.79)	3.76	(0.67)	0.53	(0.70)
PTP (*n* = 17)	3.25	(0.42)	4.00	(0.37)	0.74	(0.52)
Total (n = 53)	*3.25*	*(0.62)*	*3.88*	*(0.57)*	*0.62*	*(0.62)*

[a] The scale maximum is 5. The higher the score, the higher the appreciation.

4.3 Motivation: Self-efficacy

A repeated measures ANCOVA was conducted with condition as the independent measure, gender as covariate and the two *self-efficacy* measures as dependent variables. This yielded a main effect of time ($F(1,49) = 13.73$, $p = .001$). Table 3 shows that self-efficacy was enhanced in all three conditions. There was no significant effect of the covariate gender. However, there was a significant interaction for time and condition, $F(2,49) = 4.724$, $p = .013$. This result indicates that self-efficacy change differs across conditions. Pairwise comparisons showed that the self-efficacy change tended to differ between PT and PTP ($p = .076$, $d = 0.61$). More importantly, the conditions TP and PT differed significantly ($p = .010$), with a very high effect score ($d = 1.13$). This finding supports the earlier finding that the TP-condition is the strongest condition on self-efficacy.

Table 3

Means for Self-efficacy by Condition

	Self-efficacy					
	Before[a]		After[a]		Growth	
Condition	*M*	(*SD*)	*M*	(*SD*)	*M*	(*SD*)
TP (*n* =18)	2.67	(0.89)	3.87	(0.63)	1.88	(1.09)
PT (*n* = 18)	3.41	(1.03)	3.76	(0.67)	0.61	(1.16)
PTP (*n* = 17)	3.21	(0.91)	4.00	(0.37)	1.24	(0.89)
Total (n = 53)	*3.10*	*(0.98)*	*3.88*	*(0.57)*	*1.24*	*(1.16)*

[a] The scale maximum is 5. The higher the score, the higher the appreciation.

4.4 Learning: Conceptual and procedural knowledge

The data on *conceptual knowledge* as a proxy for learning shows that participants scored high on recalling conceptual information about the learned tasks (see Table 4). There was no significant difference between conditions, $F(2,50)$: 2.068, $p = .137$. It must be noted that there was a main effect of gender on conceptual knowledge (($F(2,50)$: 4.206, $p = .046$). When looking at condition pairs, this covariate was significant only for the PT- and PTP condition pair ($p = .018$). However, the Cohen's *d* here is 0.11, which is perceived as a low effect.

Table 4

Means for Knowledge[a] by Condition

	Knowledge			
	Conceptual		Procedural	
Condition	*M*	(*SD*)	*M*	(*SD*)
TP (*n* =18)	5.17	(0.98)	5.83	(0.38)
PT (*n* = 18)	4.78	(0.87)	5.06	(0.80)
PTP (*n* = 17)	4.88	(0.85)	5.59	(0.87)
Total (n = 53)	*4.94*	*(0.90)*	*5.49*	*(0.77)*

[a] The scale maximum is 6. The higher the score, the more items correct.

Turning now to the data on *procedural knowledge* as a proxy for learning, the participants scored high on completing the hands-on tasks during the test. There was a main effect of condition on procedural learning, showing significant difference between conditions (corrected for gender, $F(2,50)$: 4.079, $p = .023$, $d = 2$). There was no effect of gender. Pairwise comparison showed an interaction effect and trend. What stands out is the TP-PT-condition pair. These conditions differ significantly from each other ($p = .005$) with a very high Cohen's *d* of 1.22. Also noteworthy is that the PT-PTP-condition pair shows an 'almost' significant difference ($p = .053$) with a high effect score ($d = 0.80$). This indicates that PT is the 'worst' condition.

5. Discussion

Initial literature research indicated a potential of timing of practice on learning and motivation but failed to present solid empirical support. The aim of the current study was to bridge this gap by determining the effect of timing of practice on engagement, motivation and learning. This was examined by an experiment that sequenced tutorials and practice. The following paragraphs will discuss and conclude the findings of this study.

5.1 Engagement

The first research question investigated the effect of the timing of practice on engagement. No significant effect was found. Nonetheless, the percentage of the videos that has been watched was very high. On average, 96.79% of the videos was played at least once. This suggests that these were videos of notable quality as the videos retained the interest of the participants, which is likely to have increased motivation (Bandura, 1977). Overall, the study findings underline that the tutorials did improve motivation and learning, as their design was strong and scientifically valid.

The solid video design may have been a factor in the fact that video segments were rarely replayed. As replay hardly occurred, there was refrained from further analysis.

5.2 Motivation: task relevance and self-efficacy

The second research question investigated the effect of timing of practice on motivation. Overall, the results showed positive change in task relevance scores and self-efficacy scores as a proxy for motivation. An explanation for this is that the design of the tutorials strengthened task relevance and self-efficacy development by following several theoretical models and guidelines, such as Brar and van der Meij (2017). For example, the inclusion of anchoring in the task domain (M1) and arranging practice before a tutorial (P2) is likely to have supported task relevance development. The inclusion of segmentation (R1), pausing (R4) and reviewing (R5) as retention boosters and the (arrangement of) practice (P1, P2) and practice files (P3) as production boosters are likely to have supported self-efficacy development.

***Task relevance*.** Regardless of tutorial design efforts, the positive change in task relevance scores was not significant. Timing of practice did not significantly influence task relevance. Other research regarding (the timing of) practice, did not find any implications of practice influencing task relevance as well (e.g. van der Meij, 2018; van der Meij et al., 2017; van der Meij & Dunkel, 2020). It is unclear why no significance was found, but it is possible that initial task relevance was already high, prohibiting major changes in scores.

***Self-efficacy*.** Contrary to task relevance, significance was found among the results regarding self-efficacy, resulting in two main findings. The first finding is that all group scores advanced significantly, indicating a steady and successful design of the training activities. Such an effect is also found in the practice in software training study of van der Meij and Dunkel (2020). This is different from the results of van der Meij (2018) and van der Meij et al. (2017), where no significance was found despite the application of similar theories and guidelines to the tutorial designs.

An explanation for this contradiction is likely to be the availability of extrinsic feedback, which was applied to the current study as an instructional feature, likewise to the study of van der Meij and Dunkel (2020). Over the past decades, research repeatedly found feedback to be effective on increasing self-efficacy (e.g. Bandura, 1977; van der Meij & Dunkel, 2020), and integrating extrinsic feedback in learning situations is therefore recommended by Fiorella and Mayer (2019).

In earlier studies regarding (the timing of) practice, such as van der Meij (2018) and van der Meij et al. (2017), extrinsic feedback was not available. Perhaps unsurprisingly, it is these studies that did not find significant differences among conditions on self-efficacy after the training and self-efficacy growth. It can be suggested that these studies assumed that the internal feedback that software trainings naturally provide would be enough to facilitate self-efficacy gains (van der Meij & Dunkel, 2020). That is, the human-computer-interaction between the user and the interface sustains task progression, as the realistic usage of an interface does not require the learner to recall entire sequences of steps, but merely requires the learner to recall how to move on from any point in this interface (van der Meij & Dunkel, 2020).

The second finding is that the timing of practice had a significant effect on self-efficacy scores after training and on self-efficacy change. This indicates that the learner's perceived capabilities regarding the task, to some extent, depend on the moment practice occurs around the video tutorial. Looking into this more deeply, it came as a surprise that the PTP-condition was not the strongest on self-efficacy scores and change. It was expected that the more practice a learner experiences, the higher one's self-efficacy would be. The found results do not support this. More striking perhaps is that the TP-condition increased significantly more strongly compared to the PT-condition. This PT-condition scored weakest in both self-efficacy after the training and self-efficacy change. A suggestion can be made that the classic coupling of practice following training remains advantageous.

A limitation of the current study regarding self-efficacy is perhaps a misinterpretation of the initial self-efficacy item. That is, the initial item asked the learner's current self-efficacy state ("how well do you think you can complete this task"), but some respondents may have answered it as a future state (e.g. the thought "I will learn it, so I will be able to do well").

5.3 Learning: conceptual and procedural knowledge

The third research question investigated the effect of the timing of practice on learning. Here, a distinction was made between conceptual and procedural knowledge, which were measured immediately after training.

***Conceptual knowledge*.** The investigation of the effects of timing of practice on conceptual knowledge has not been done in other practice studies (e.g. van der Meij et al. 2017; van der Meij 2018; van der Meij & Dunkel, 2020). Overall, students achieved high scores on the conceptual knowledge test. This indicates that the videos and practice activities were efficient on learning. Nonetheless, there was no significant difference across conditions. A possible explanation is that the practice activities focused more on procedural knowledge than on the acquisition and rehearsing of conceptual knowledge. Due to the novelty of integrating conceptual knowledge in research regarding the timing of practice, future research should investigate the influence of (the timing of) practice on conceptual knowledge.

***Procedural knowledge*.** Regarding procedural knowledge, scores were very high on the test for all conditions. These high learning scores indicate that the optimization of used videos and practice activities was effective.

Not only did all conditions score high, they also differed significantly from each other, which was likewise found in van der Meij and Dunkel (2020). In the current study, the TP-condition was the strongest scoring condition and the PT-condition the weakest, whereas the PTP-condition was located somewhat in the middle. When looking at condition pairs, the TP-condition even significantly differed from the PT-condition, with a very large effect score. Another result of the pairwise comparison is that the PTP-condition especially favorably compares to the PT-condition, with a large effect score, with an almost significant *p*-score. Both findings of significant imbalance between conditions contradict the initial expectation, that predicted that the PTP-condition would outperform the TP- and PT-conditions because of the combination of their strengths. It remains unclear as to why the conditions differ (significantly) from one another. Future research is needed to provide explanations for these differences. What can be said is that the classic coupling of practice following the tutorial appears to remain favorable, as was the same case regarding the findings on self-efficacy.

Furthermore, the current findings on procedural knowledge contradict the findings of van der Meij (2018) and van der Meij et al. (2017), in which no significant effect of timing of practice was found. An explanation for this contradiction can be found in van der Meij and Dunkel (2020), as they state that significant influence of timing of practice depends on the accessibility of extrinsic feedback.

5.4 Conclusion

The current study investigated the influence of timing of practice on engagement, motivation and learning. The study did this through the alteration of practice activities and viewing activities, both regarding DBT-videos. The videos of notable quality since they were optimized for fostering learning based on design guidelines and theories.

In conclusion, it does not seem that the timing of practice is effective on engagement, task relevance and conceptual knowledge. However, the timing of practice does influence self-efficacy and procedural knowledge. These found effects can be attributed to the availability of extrinsic feedback. One of the more remarkable findings to emerge from this study is that the coupling sequence of practice after instruction, that already is well-established in education, remains favorable.

However, a limitation of the current study is that it is unclear what exactly happened during practice activities. For example, no data is available on how long participants practiced for or how cognitively engaged they were during such activities. Future systematic research should take a detailed look into learner’s behavior during practice. The current study is a step towards an overview of how practice as an instructional feature can function optimally in software training. In a broader sense, future studies like the current one will help develop guidelines on how instructional features can be used optimally in relation to DBT-tutorials. Ultimately, more research will contribute to establishing an overview of what instructional features are effective under what circumstances, so that people’s learning can be maximized.

References

Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice-Hall.

Brar, J., & van der Meij, H. (2017). Complex software training: Harnessing and optimizing video instruction. *Computers in human behavior*, *70*, 475-485.

Clark, R. C., & Mayer, R. E. (2003). *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning*. John Wiley & Sons.

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. Dissertationsschrift, Albert-Ludwigs-Universität Freiburg*.

Fiorella, L., & Mayer, R.E. (2018). What works and doesn't work with instructional video. *Computers in Human Behavior*, *89*, 465-470.

Grabler, F., Agrawala, M., Li, Wi., Dontcheva, M., & Igarashi, T. (2009). Generating photo manipulation tutorials by demonstration. *ACM Transactions on Graphics, 28*(3), 66.11-66.69.

Grossman, R., Salas, E., Pavlas, D., & Rosen, M.A. (2013). Using instructional features to enhance demonstration-based training in management education. *Academy of Management Learning and Education*, *12*(2), 219–243.

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

Kay, R. & Kletskin, I. (2012). Evaluating the use of problem-based video podcasts to teach mathematics in higher education. *Computers and education, 59*, 619-627.

Kosslyn, S. M., Kievit, R. A., Russell, A. G., & Shephard, J. M. (2012). PowerPoint® presentation flaws and failures: a psychological analysis. *Frontiers in psychology*, *3*, 230.

Koumi, J. (2003). Synergy between audio commentary and visuals in multimedia packages. *Journal of Educational Media, 28(*1), 19-34.

Mayer, R.E., Heiser, J. & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology, 93*(1), 187-198.

Mayer, R.E. (2008). Applying the science of learning: evidence-based principles for the design of multimedia instruction*. American Psychologist, 63*(8), 760-769.

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.

Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and instruction*, *2*(1), 59-89.

Spanjers, I. A., van Gog, T., Wouters, P., & van Merriënboer, J. J. (2012). Explaining the segmentation effect in learning from animations: The role of pausing and temporal cueing. *Computers & Education*, *59*(2), 274-280.

van Gog, T. (2011). Effects of identical example–problem and problem–example pairs on learning. *Computers & Education*, *57*(2), 1775-1779.

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*(3), 212-218.

van der Meij, H. (2018). Cognitive and motivational effects of practice with videos for software training. *Technical communication*, *65*(3), 265-279.

van der Meij, H., & Dunkel, P. (2020). Effects of a review video and practice in video-based statistics training. *Computers & Education*, *143*, 103665.

van der Meij, H., & van der Meij, J. (2013). Eight guidelines for the design of instructional videos for software training. *Technical Communication, 60*(3), 205-228.

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

van der Meij, H., & van der Meij, J. (2016). The effects of reviews in video tutorials. *Journal of Computer Assisted Learning*, *32*(4), 332-344.

van der Meij, H., van der Meij, J., Voerman, T.E., & Duipmans, E. (2017). Supporting motivation, task performance and retention in video tutorials for software training. *Educational Technology Research and Development*, 1-18.

van der Meij, H., Rensink, I., & van der Meij, J. (2017). Effects of practice with videos for software training. *Computers in Human Behavior*, 1-7.

van der Meij, J. & van der Meij, H. (2014). A test of the design of a video tutorial for software training. *Journal of Computer Assisted Learning*, 1–17.

Wigfield, A., & Eccles, J. S. (2000). Expectancy–value theory of achievement motivation. *Contemporary Educational Psychology*, *25*(1), 68-81.