Monitoring and Interactivity in Instructional Video

Thilo Doepel

Faculty of Behavioral Sciences

Department of Instructional Technology (IST)

Examination Committee:

First Supervisor: Dr. Hans van der Meij

Second Supervisor: Dr. Hannie Gijlers

Master Thesis, June 2014

University of Twente, Enschede

UNIVERSITY OF TWENTE.

Abstract

This study examines the effectiveness and influence of a self-directed-learning (SDL) environment used for facilitating the learning of primary school students (aged 10-12). An interactive instructional video is used in combination with the assessment of meta-cognitive monitoring ability through the Feeling of Knowing (FOK) effect. The underlying rationale is that video may induce high cognitive load, which in turn could be reduced through FOK assessment and interactive use of the video. 110 participants from five classes are randomly assigned to one of four conditions. The control group receives the regular video instruction without the support of the two effects in question. The instructional video falls into the domain of formatting skills within Microsoft Word 2010. Pretest, training score, posttest, and retention test data are collected from all the groups. Measurements include: learning gains (a measure of cognitive performance), a motivational questionnaire, an assessment of monitoring ability and examining the use of interactivity. The findings favored no condition significantly over the other, showing that neither the assessment of the metacognitive ability nor the interactive components supported the learners effectively. The discussion leads us to believe that monitoring abilities like FOK should be integrated into SDL environments very carefully and that video tutorials can successfully support procedural learning.

Keywords Dynamic Representations · Feeling of Knowing · Instructional Video · Interactivity · Monitoring · Procedural support · Self-directed-learning

Deze studie onderzocht de effectiviteit en de invloed van een zelfstudie leeromgeving (*self-directed-learning* environment) op het leren van basisschool leerlingen tussen de 10 en 12 jaar oud. Een interactieve instructie video wordt gebruikt in combinatie met de vraagstelling van een metacognitieve trigger, het *Feeling of Knowing* effect (FOK). De onderliggende gedachte is dat video instructie cognitief inspannend is voor de scholieren en dat dit gereduceerd kan worden door FOK taken en de interactieve functies van een video. Het wordt verwacht dat beide variabelen effect hebben op het succes van de kinderen met de opgaven. 110 scholieren uit vijf klassen zijn random over één van vier condities verdeeld. Erbij krijgt de controle groep alleen de video instructie zonder ondersteuning van het FOK effect en de interactieve functies. De gebruikte instructie video valt in de categorie van computer vaardigheden omdat de leerlingen leren hoe ze beter met de tekst verwerking functies binnen Microsoft Word 2010 om kunnen gaan. De cognitieve prestatie wordt tijdens de studie vier keer gemeten: tijdens een Voorkennis test, gedurende de Training, bij een Natest en tot slot bij de Retentie test. Daarnaast worden er nog een motivatie vragenlijst alsmede twee metingen voor het FOK effect en de interactiviteit afgenomen. De resultaten tonen geen effect van conditie aan en daarom wordt bij deze studie aangenomen dat het reflecteren over het eigen leren en de interactieve functies de leerling niet significant ondersteunen. Binnen de discussie wordt vanuit de literatuur duidelijk dat de twee effecten het leer succes wel zouden kunnen verhogen als ze voorzichtig in de leer omgeving geïntegreerd worden.

Table of Contents

Introduction

Background

The aim of this study is to explore the effects of interactivity within procedural video instruction and the Feeling of Knowing phenomenon (FOK) and its accuracy on children's learning gain and behavior with an interactive instructional video within a self-directed-learning (SDL) environment. In order to make these concepts understandable and to make their connections clear the key terms are explained and put into context in the following sections. The connection of FOK phenomenon, as a part of the metacognitive monitoring ability, and its influence on the self-directed-learning experience with the interactive functions of an instructional video as a learning tool is laid out. In this case instructional videos are the kind of *How to* – videos found all over the internet in which the steps of a particular skill are modeled for the viewer.

Self-directed-learning

The use of the self-directed-learning method has become a common practice in modern classroom settings. Teachers try to implement it into the curriculum in order to give the learner control of his own learning processes.

There are four dimensions to self-directed-learning: Personal autonomy, self-management, learner control and the independent pursuit of learning (Candy, 1991). Knowles (1975), one of the first researchers of the SDL approach, defined it as a process by which individuals take the initiative, with or without the assistance of others, in diagnosing their learning needs, formulating learning goals, identifying human and material resources for learning, choosing and implementing appropriate learning strategies, and evaluating learning outcomes.

SDL can be introduced to the classroom with the help of open programs, individual study options, and non-traditional learning methods. It is considered a learning characteristic in every human being and involves activities like self-guided reading, electronic dialogues, individualized study options, working with hypermedia (Azevedo & Cromley, 2004; Blumberg, 2000), participation in study groups or learning through internships. All those activities empower the learner to take more responsibility for all the activities of the learning endeavor (Hiemstra, 1994). In the classroom, the teacher can play a supportive role wherein he scaffolds the pupils' SDL through explaining worked-out examples, handing out materials and helping the learner when he gets stuck in his learning process. The teacher can also model learning strategies and help the learners to make them their own, thereby making learning more "visible" (Corno, 1992).

SDL can include domain-specific learning environments, which help the learner to focus on problems they have in the particular domain and provide contextualized support through specific tasks (Fischer & Scharff, 1998). This also often provides them with the ability to transfer conceptual knowledge to new situations. SDL is often seen as learning that comes close to learning from real-world problems by considering how people learn in real life (Bolhuis, 1996; Temple & Rodero, 1995). The advantages of SDL include the way it allows the learner to learn at his own pace and receive feedback about his performance and the better and direct access to training material. The most influential and common disadvantage, especially for younger learners, stems from the fact that the learner must have the ability to learn on his own.

Several studies showed that the use of instructional videos can help students to develop a more independent approach to learning (Biggs & Tang, 2007; Luke, 2011). The independent learning, forming an important aspect of SDL, can be supported by the videos as they form an educational system in which the learner is autonomous and separated from his teacher so that communication is by print, electronic, or other non-human medium (Moore, 1973).

Unlike traditional learning approaches, SDL requires learners to engage in their own metacognitive processes to monitor and adjust their learning strategies (Bolhuis, 1996; Garrison, 1997; Zimmerman and Martinez-Pons, 1990). The metacognitive processes that can occur within SDL can facilitate the development of creative and critical thinking (McCombs and Vakili, 2005). It encourages students to evaluate themselves and their learning outcomes. The learner checks his own learning progress and decides on how to approach the task, thereby managing the whole learning experience (Morrow, Sharkey, & Firestone, 1993). The learners who become aware of the strengths and weaknesses of their knowledge, strategies, affect and motivation, are better able to regulate their own learning by controlling and monitoring the cognitive processes needed (Peterson et al., 1982).

Within the Cognitive-Affective Theory of Learning with Media (CATLM) the role of metacognitive processes as mediators for learning is pointed out (Moreno, 2006, 2009). The theory discusses the regulatory role metacognitive thought processes can take on when the learner is confronted with new multimedia tasks, which he must solve independently in a SDL environment.

The way metacognitive self-monitoring, with a focus on FOK, is essential for SDL and its effectiveness, as pointed out by Bolhuis (1996), is explained in the following section.

Self-Monitoring

The complex phenomenon of metacognition refers to the cognitive control and monitoring of many sorts of cognitive processes like perception, action, memory, reasoning or emoting. There are several descriptions of metacognition and its working: Flavell (1979) describes it as “cognition about cognitive phenomena” or “thinking about thinking” while Martinez (2006) regards it as “the monitoring and control of thought". While metacognition knows many processes and effects, only self-monitoring and the Feeling of Knowing effect (FOK) are discussed here.

Self-monitoring is part of one's ability to evaluate the own comprehension and understanding of subject matter and use that evaluation to predict how well one might perform on a task. This forms a very important aspect for SDL. The learner who has an advanced awareness of his knowledge states through monitoring, is in a better position to take action for his learning and to change his actions or behavior when his initial learning strategy does not work out, than an unaware learner (Peterson et al., 1982; Thiede, et al., 2003). A learner is using a part of his monitoring ability if he realizes that he is having more trouble learning how to complete a fraction problem than a multiplication problem.

A theoretical framework of metacognition that can be used to put the workings of self-monitoring into context is the Good Information Processing Model by Pressley, Borkowski and Schneider (1989). This model shows how metacognitive knowledge is related to the learner's strategy use, motivational orientation, general knowledge about the world, and automated use of efficient learning procedures, not only for conceptual knowledge but also for procedural and declarative knowledge. The components of this model are assumed to interact and have an effect on knowledge acquisition.

Self-monitoring also fits into Swartz and Perkins (1989) fourth level of metacognition, the reflective use, as the individual reflects upon his or her thinking before, after or even in the middle of the process, evaluating how to proceed and how to improve (see also Livingston, 1977).

In order to put everything into context it is important to note that the metacognitive monitoring process is one of the areas of metacognition that improves with age, as older children are better than younger children at monitoring abilities like FOK. This improvement can be supported through instruction programs at schools (Schneider and Lockl, 2002; Schneider, 2008). It is important to address the monitoring process of FOK specifically in this context, because SDL leaves the learner in charge of the evaluation of his own knowledge gain and through the FOK the learner can monitor if the desired knowledge can be recalled from memory.

The subjective feelings of FOK judgments and other metacognitive monitoring phenomena seem to guide and affect our behavior before, during and after learning processes, but it is still unclear how they do it exactly, which factors play a role and how we can influence them (Koriat and Goldsmith, 1996; Nelson, 1990). The subject of FOK was pioneered by Hart (1965; 1967) through his research on the empirical assessment of the accuracy of the FOK by presenting the subjects with general-information questions and then asking them to attempt to recall the correct answer to each question. When the participants could not recall the criterion response they were asked to make a FOK judgment, estimating whether they would recognize the non-recalled answer if given a forced-choice recognition test (Nelson, 1996).

An example of FOK is if you cannot remember the answer when someone asks you what the capital of Latvia is, but you feel that you would recognize the name if you saw it on a map of Europe. If your direct recall from memory fails and no criterion response can be retrieved you might be able to monitor your learning progress and make a FOK judgment, assessing the likelihood of your ability to recall the information at a later point in time (Brown, 1991; Schwartz, 1994). The FOK can be assessed before or after retrieval of the information has failed and can take the form of a statement like "Definitely I could recall [the information] ", whereby high and low FOK ratings are possible. The participant can monitor the knowledge in such a way that he feels that the information is located somewhere in memory but is unable to recall it at the given moment (Hart, 1965; Reder and Ritter, 1992).

The question on how the learner a FOK state knows that he ''knows'' the sought-after target in the face of being unable to produce it remains and raises interest. Especially so when considering the empirical findings that indicate that FOK judgments made after retrieval failures are moderately valid in predicting the success of retrieving the sought target or recognizing it from among distractors at some later time (Schwartz and Metcalfe, 1992). A possible explanation could include the idea that FOK states directly detect the presence and, perhaps, the strength of memory traces.

The confidence the learner has in his FOK judgments depends largely on the learner witnessing the outcome of his own controlled processing, thus it is by retrieving a solicited answer and noting the amount of effort expended in its retrieval that we form our confidence in the correctness of that answer. It can be considered that there is a feedback loop from controlled action to subjective monitoring, and perhaps more generally, from behavior to consciousness (Koriat, et al., 2001). Researching the nature of FOK judgments Schneider and Lockl (2008) found supporting evidence for Koriat's ''trace accessibility'' model (1993, 1995), stating that FOK judgments are based on retrieval attempts and determined by the amount of information that can be spontaneously generated, regardless of its correctness.

The advantage the inclusion of a FOK assessment in a SDL environment has is centered on the fact that even when it fails, the act of trying makes the pupil aware of his own understanding of the learning matter. In his comment on Schraw (1995), Wright (1996) states that FOK judgments are often measured and analyzed using the Hamann's coefficient (HC) or Goodman and Kruskal's γ. Wright supports the use of γ, as it appears to be a direct measure of the diagnostic worth of FOK ratings. A Feeling of Knowing judgment can be made about all kinds of knowledge and information. The original source of information does not interfere with the ability to think about its recallability. Even though the process appears independent from the medium, the accuracy of the FOK might.

The accuracy of the Feeling of Knowing judgments depends on the difference between the

FOK hit and FOK miss rates, which shows to be significantly smaller when the recognition alternatives depend upon the recall response than when they do not (Blake, 1973). It can be stated that most of the learner's FOK judgments are fairly valid predictors of recognition for recently presented, unrecallable information and regardless of whether partial recall is free to influence recognition, likelihood of the Feeling of Knowing response increases as a function of amount of recall of the information. There are no studies reporting results on FOK accuracy of procedural knowledge. Throughout childhood and adolescence increasing FOK accuracy with increasing age has been reported by most studies (see Schneider and Lockl, 2008), but this increase seems to come to a hold in adulthood.

By proving that monitoring processes like the FOK can play a central role in directing how people study, Schneider (2008) showed that we can expect a direct link between the results of a learner's FOK and the following behavior (Thiede and Dunlosky, 1999; Son and Metcalfe, 2000). All this information underlines the fact that the assessment of the FOK can have strong influences on the learning experience. The way in which all these concept can be tied into the learning from instructional video, will be discussed in the next section.

Instructional video

Interactive instructional video is an important element in software training and the important aspects of its workings need to be considered in order to support learning the most efficiently. As described in the section on SDL, instructional videos can be an important scaffold for overcoming the disadvantages of this learning approach. Azevedo and Cromley (2004) show in their study that learning from hypermedia is better supported by a SDL approach to learning than by a traditional learning approach. Instructional videos can easily be integrated into a SDL environment through the use of computers and/or a beamer. Furthermore, the FOK can be assessed perfectly in combination with this tool, as it takes over the evaluative part of the learning experience, often not provided in many instructional videos.

New parts of the learning environments for teaching and learning have come to light in the last two decennia and have given rise to new areas for research. Instructional video is only one of the developments that can support the learner in a world where the time-to-competency is continuously reduced (see Germany's reduction of school time within High school). The format of instructional videos is well suited for delivering an effective load of visual and verbal information, their distribution over the internet is easy and economical and their interactive use a clear advantage over non-interactive media. Schools and teachers have embraced the effectiveness instructional videos can have for their learners. In comparison to a paper-based tutorial, an instructional video proves to have a higher success rate (van der Meij and van der Meij, 2013).

Students nowadays already show a slight degree of independent learning, as shown by Kennedy and colleagues (2006), because they are familiar with using platforms like *youtube* and are able to clarify concepts discussed in class. Instructional videos are different from lecture recordings and other online resources focused on spreading knowledge, as they display the practical, real-world side of the instruction process more clearly as each aspect of the procedure is shown directly and not approximated by a graphical or figurative display. The familiarity with this kind of learning is thus increased (Luke and Hogarth, 2011). Instructional video can also deliver a stronger flow than other learning material (Vollmeyer & Rheinberg, 2006).

Mayer's Cognitive Theory of Multimedia (2005) offers insights about the way instructional video influences and supports the learner in his learning process. The distribution of information via different channels, as it is the case with a video with narration, reduces the cognitive load of the learner. The learner is encouraged to engage in processes such as information selection, organization and integration with existing knowledge. Instructional videos are used to teach the pupils a procedural skill in a clear, visual and information-rich manner. Many instructional videos concerned with the teaching of a computer skill contain a voice over narrative explaining the steps needed to be taken to master the procedure. The narrative can partially replace the dependence on a teacher, which is discouraged in a SDL environment. Children in the seventh and eighth grade classes are often introduced to software from the *Microsoft Office* collection as it can help them with their schoolwork and future academic endeavors.

While the importance and general effectiveness of instructional video for learning is not easily questioned, the quality of each individual instructional video is not guaranteed to reach the full effectiveness of the medium. Mayer and Moreno are two of the leading researchers in this field of study and have published several studies on the workings of interactive multimodal learning environments (2003; 2007).

The design of good videos can prove difficult and requires careful consideration of design guidelines so that the material presented is manageable, broken down into appropriate segments (Zheng, et al., 2009), linked to the right learning outcomes (Race, 2005) and has a focus on key processes and steps (Harp and Mayer, 1998). Next to those elementary ideas there are is a set of eight guidelines for the construction of a good instructional video that has proven useful (van der Meij, 2013). Using these guidelines has been proven to improve the way the users find, perceive, understand and recall the given information within and after use of the video, thus improving the entire task performance (van der Meij, 2013).

In order to let the students retain enough cognitive capacity to concentrate on deep processing it is important to maximize the effort that learners put into elaborating content while minimizing the effort they must expend to make sense of this content (Cennamo, 1994), which can be achieved by setting up the video according to design guidelines.

Ertelt (2007) shows in her doctoral dissertation that the guidelines for modeling a worked-out example and Labelling are effective tools and that Practice and Pacing should be included to ensure general procedural knowledge acquisition. Especially the most recent guideline, self-pacing of the video content by the learner, is within the center of this study. Self-pacing has been demonstrated to facilitate learning with the medium (Mayer and Chandler, 2001). The functions of stopping, starting and replaying an animation can allow re-inspection, focusing on specific parts and actions. Next to that the animations that allow close-ups, zooming, alternative perspectives, and control of speed are likely to facilitate perception and comprehension (Tversky, et al., 2002). Mayer and Chandler (2001) showed that students who had control over the pace of the presentation through the simple means of a Continue-button performed better on subsequent tests of problem-solving transfer than those who received a continuous presentation.

Using this information one can see that instructional video can be an effective tool for SDL. Within the next section it is outlined how the learning from instructional video and the assessment of FOK are combined in this study.

Ideas behind this study

The implications that can be drawn from this literature review form the basis of the ideas behind this particular study. With the focus on the workings of FOK assessment, Self-pacing options and their interaction within learning from an instructional video, we designed the following concept for a study.

The subjects are placed into a SDL environment with an instructional video as the learning tool and the assessment of the FOK as an added dependent variable. Instructional video and FOK have both been proven effective means to support the workings of SDL (Tversky, et al., 2002; Perrotin, Tournelle and Isingrini, 2008).

As Nelson (1996) showed that the learner's self-monitoring can give important clues about what people know about themselves and why they choose to behave as they do, we decided to measure the learners' behavior with the tool through the use of a questionnaire and observation in a training session, wherein their FOK is assessed several times. Because we assume that the FOK assessment influences the pausing, playing, skipping forwards and backwards and the redoing of a practice segment of the video, meaning that the learner's behavior is influenced, we choose the concept of self-pacing of the video as the other dependent variable. This assumption stems from the idea that when the learner cannot recall the newly acquired knowledge during a FOK assessment he might consult the video again for further information.

We do not expect the FOK assessment to have strong influence on van der Meij's other design guidelines. Furthermore we test the effects on learning gain from the instructional video and the FOK assessment per pupil by examining scores from pre-test, training, post-test and retention-test. The influence of differently designed instructional videos on the learner's behavior by provoking different FOK states is within the focus of this study.

Within the seventh and eighth year of school young learners are first introduced to learning procedural knowledge from instructional video. Most of them are expected to have experience with platforms like *Youtube* and the text processer *Microsoft Word*, but not in an academic and computer context.

It is expected that the FOK effect will behave differently with learned procedural knowledge than with simple answers from memory traces, like consonant trigrams. The information is more complex but also can inherit a more familiar real-life feel for the learning experience, which might ease the acquisition of new knowledge. The FOK accuracy is assessed on several points across the experiment through letting the participants estimate their ability to recall of parts of the learned procedure in the form of the statements. It will be measured through the FOK hit rate and FOK misses in combination with the post-test scores. We included this measurement in order to see if the children got a good grasp on their own knowledge and learning strategies. The four different conditions (shown in the table below), in which this study is split in, are differentiated according to the application of interactivity (on the level of self-pacing in this study) for the instructional video and FOK states assessed:

Learner trains with an **interactive** video and answers **metacognitive** questions (Cond.1)	Learner trains with an **non-interactive** video and answers **metacognitive** questions (Cond. 3)
Learner trains with an **interactive** video and **does not** answer **metacognitive** questions (Cond. 2)	Learner trains with an **non-interactive** video and **does not** answer **metacognitive** questions (Cond. 4)

Prior knowledge (cue familiarity and general familiarity with computer tasks) and the use self-pacing are expected to be most influential factors for high ability to recall of the criterion responses. The video confronts the children with computer problems and SDL is the correct learning approach as it encourages them to use their ability to problem-solve (Bolhuis, 1996). Procedural knowledge, being the knowledge of how to use a selected strategy, is the kind of knowledge that needs to be acquired to solve the computer problems. We also hope that by assessing FOK several times during the training with the instructional video the children might get used to the concept and apply in their learning environment outside of the training, as some studies show that learning about metacognitive strategies induces increases in learning (Vermunt and Vermetten, 2004).

Research questions and hypotheses

The research questions resulting from the theoretical background and the plans of adapting it within this study are:

1. Does the video tutorial support procedural knowledge acquisition?
2. How does the assessment of FOK states affect the learners' learning gain from the instructional video?
3. How does the application of self-pacing options affect the learners' learning gain from the instructional video?
4. Is there a difference in FOK assessment between the age groups?
5. Which measures predict task performance?

The hypotheses tested under these conditions are: The video tutorial supports the procedural knowledge acquisition and the design guidelines appear to have formed a successful video tutorial. Prior knowledge and the successful application of van der Meij's (2013) self-pacing guideline for instructional video increase the learner's task performance and the FOK hit rate. Furthermore the assessment of the FOK state increases the learning gain showed on the retention-test, as the metacognitive processes used to reflect about one's own knowledge might not only influence the monitoring of knowledge but also the actual learning transfer through the behavior with the video. The students who reflect about their learning and can also pace the video tutorials according to their own needs are expected to show the best task performance.

The hypotheses handled result in an experimental study with a 2 (self-pacing, no self-pacing) × 2 (FOK questions, no FOK questions) between-subjects factorial design see the table on the conditions), with the only independent factor being condition. The dependent measures were the results of the IEMQ-questionnaire (Motivation), FOK assessment, the use of the self-pacing functions and the four measures of task performance.

Method

Participants

A total of N=115 participants (65 female and 50 male) was approached for this study. The participants came from five classes of three primary schools from a suburb of Enschede, Netherlands. There were 54 students from the seventh grade (29 girls) and 61 students from the eighth grade (36 girls) ranging between 10 and 13 years of age and with an overall mean age of 11 years and 6 months (SD = 0.8). Students were randomly assigned to one of the four conditions. According to the teachers the students from that particular suburb stem from a mainly low SES (Socioeconomic status) background which was reflected in the learning behavior of some students (low motivation and small interest in education) and contributed to the fact that some had very little experience with computers. This facet should be taken into account when comparing this study to similar ones. Through the schools written consent was received from the parents prior to the study. Due to illness during training phase, five participants were excluded from data analysis, resulting in a total number of 110 participants (49 males and 61 females). For the initial analysis of the data there were ultimately 27 participants in condition 1 (FOK + Self-pacing), 28 in condition 2 (No FOK + Self-pacing), 28 in condition 3 (FOK + No Self-pacing) and 27 in condition 4 (No FOK + No Self-pacing).

Materials

Video tutorials

The instructional material used within this study was a video whose design and content stem from a study by van der Meij & van der Meij (2014). Paper-based tutorials were compared to video tutorials for software learning with Microsoft Word. The video for this study was adapted for Microsoft Word 2010. Van der Meij & van der Meij found in earlier studies that their guidelines for the design of instructional videos for software training, served as a valuable basis for the construction of instructional video (van der Meij & van der Meij, 2014) and that the video tutorial helps students learn better methods for formatting tasks for school reports and the like (van der Meij & van der Meij, 2012, 2013).

These dynamic representations cover different tasks within Word and give the learner easy to follow instructions. The formatting exercises are organized into three “chapters”: (a) adjust the left and right margin for the whole document, (b) format the paragraphs, citations and lists and (c) create an automatically generated table of contents. The video was presented via a website. On the left side, the participants were able to click on the corresponding heading, see Figure 1, in order to call up the video tutorial, which then appeared on the right side. Visual cues, such as the one depicted in Figure 2, supported the learner’s understanding of the video tutorial. The total length of the tutorials ranged from 60 to 100 seconds. Two versions of the video were used for this study. In the interactive version the students had to select one of the headlines on the left (see Figure 1). Clicking on it displayed the video on the right part of the screen and the students could start it via the button. They were encouraged to pace the video according to their own needs of repetition with the interface at the bottom. The auto-start version of the video let it appear and immediately start automatically as soon as the students clicked on one of the headlines. The participants were discouraged from pacing each video segment by themselves and told to let it run without interfering.

Next to the visual instruction, the video contains a Dutch female voice-over, as advised by the second guideline (van der Meij & van der Meij, 2013). The narrator explains the steps that need to be taken in order to accomplish the formatting procedure successfully.

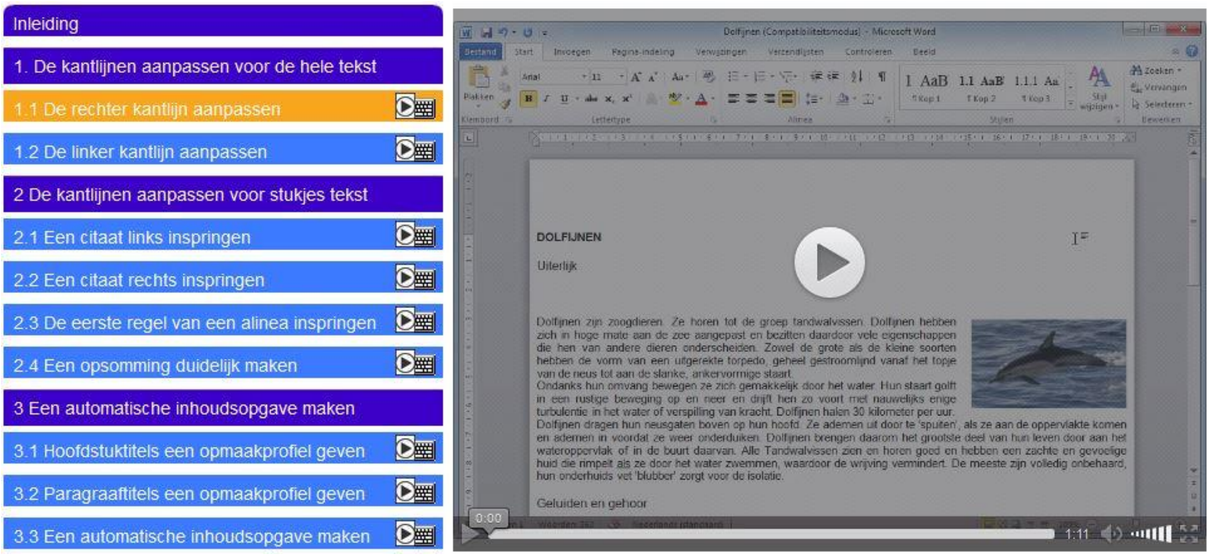


Figure 1. Website with the interactive table of contents for the activation of the tutorials

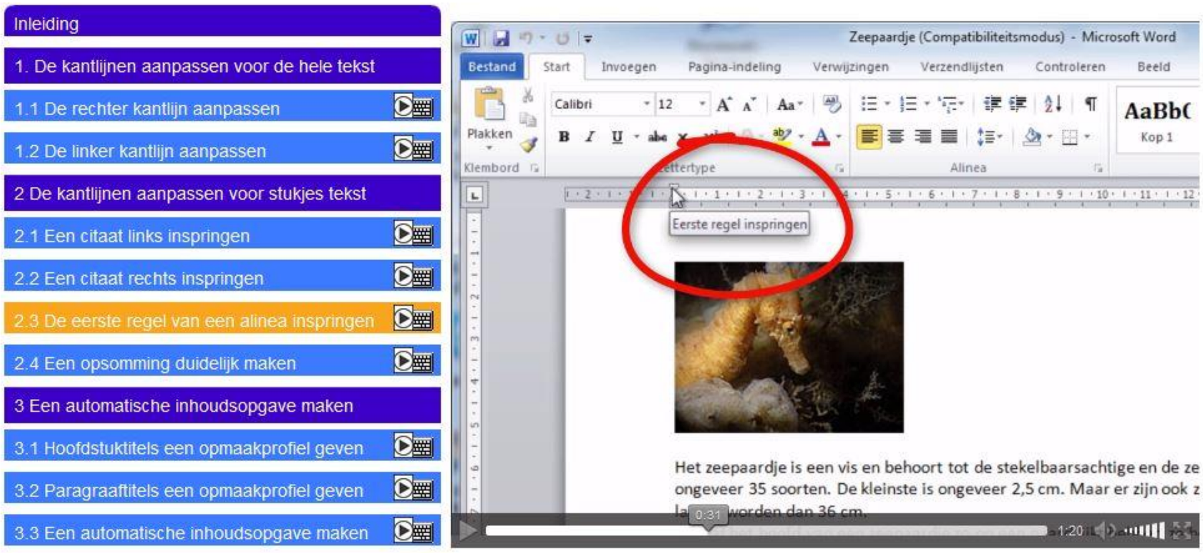


Figure 2. Visual cue for learner support

Job-aid, Practice files and Motivation measure

Next to training, there was a pre-, post- and retention test. The learners were asked to perform formatting tasks within several practice files during these four phases of the experiment and they were instructed through a job-aid each time. The job-aid always prompted the students to modify the format of practice files for the same kind of tasks that are discussed during the video tutorial. These files made task completion efforts comparable across students.

The pre-test, used to measure the students' initial task performance, consisted of five practice files and a job-aid which included an *Initial Experience and Motivation Questionnaire* (IEMQ).

The IEMQ measured the students' initial motivation. A "before" and "after" screenshot and an explanation was shown to each learner. After that the learners had to fill in three questions, before they tried replicating the procedure. Each instance consisted of three questions, whereby the first inspected the learner's previous experience with the task, the second the relevance of the task for the learner and the third the learner's self-efficacy to replicate the procedure by asking "How well do you think you can complete this task?". The answer was always scored on a 7-point-Likert-scale thus resulting in a total of 3*6 answers addressing the student's motivation.

The post-test and the retention-test asked the student to apply their recently acquired procedural knowledge. Both times the job-aid gave instructions to change one poorly formatted Word file, the given practice file, into a well-formatted exemplar. Scoring was identical across all task performance measurements (see *Codebook*).

During training five practice files needed to be handled by each student, but depending on the condition the job-aid included instructions for some students to answer the metacognitive FOK questions and/or encouraged them to use the self-pacing options.

Feeling of Knowing measure

The FOK question appeared on five moments in training. Always directly after the last video segment of one of the five main topics, the students were presented with the question "How big do you estimate your chances that you can replicate what you just learned?". The students were asked to select one of four answers: "Certain to be able to replicate the procedure", "Almost certain to be able to replicate the procedure", "Maybe able to replicate the procedure" and "Certain not to be able to replicate the procedure", receiving 3, 2, 1, or 0 points respectively.

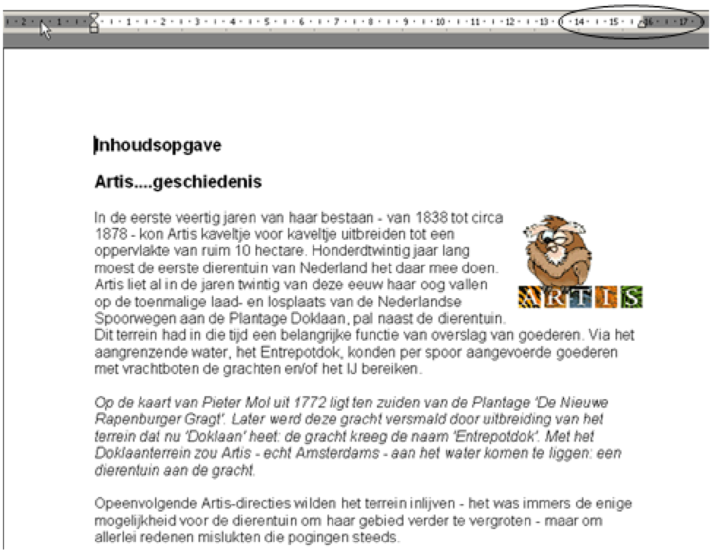
Self-pacing measure

Two questions that addressed the use of the self-pacing options in terms of quantity and efficiency, were filled in by the students immediately after training. The first question, "How often did you replay the segments or skip back and forth within the video?", concerned the quantity of the learner's use of the self-pacing options. The answers ranged from 0 to 9 or more times replayed. The second question, "Did replaying the segment improve your understanding?", could be answered by ticking "yes", "no" or "I don't know" in order to show how much the learner felt supported by the self-pacing.

Codebook

The students' task performance success was determined with the help of a codebook. The codebook gave instructions to check whether the learner replicated the procedure perfectly, reached the goal through a look-a-like solution without the correct method, or did not replicate the procedure at all.

The possible changes the learners could make in the Word documents in order to replicate the procedures were coded with 0, 1 and 2 points. They received 2 points if they reached a visually acceptable result and used the correct method to do so. 1 point was granted if they only got a look-a-like result but used a different method. Figures 3 and 4 show the difference between the right and a wrong method applied with the ruler function. No points were given if neither the right method was applied nor an acceptable result was reached. The learners could get a maximal score of 16 points in the pretest and 18 points in the other three measures of task performance. In case the learner forgot to save the changes in one of the documents the corresponding fields were left empty within the coding table. For the data analysis the missing values were replaced with the corresponding series mean and a sum-score was generated for each measure.



Inhoudsopgave

Artis....geschiedenis

In de eerste veertig jaren van haar bestaan - van 1838 tot circa 1878 - kon Artis kaveltje voor kaveltje uitbreiden tot een oppervlakte van ruim 10 hectare. Honderdtwintig jaar lang moest de eerste dierentuin van Nederland het daar mee doen. Artis liet al in de jaren twintig van deze eeuw haar oog vallen op de toenmalige laad- en losplaats van de Nederlandse Spoorwegen aan de Plantage Doklaan, pal naast de dierentuin. Dit terrein had in die tijd een belangrijke functie van overslag van goederen. Via het aangrenzende water, het Entrepotdok, konden per spoor aangevoerde goederen met vrachtboten de grachten en/of het IJ bereiken.

Op de kaart van Pieter Mol uit 1772 ligt ten zuiden van de Plantage 'De Nieuwe Rapenburger Gragt'. Later werd deze gracht versmald door uitbreiding van het terrein dat nu 'Doklaan' heet: de gracht kreeg de naam 'Entrepotdok'. Met het Doklaanterrein zou Artis - echt Amsterdams - aan het water komen te liggen: een dierentuin aan de gracht.

Opeenvolgende Artis-directies wilden het terrein inlijven - het was immers de enige mogelijkheid voor de dierentuin om haar gebied verder te vergroten - maar om allerlei redenen mislukten die pogingen steeds.

Figure 3. Correct method & result

Inhoudsopgave

Artis....geschiedenis

In de eerste veertig jaren van haar bestaan - van 1838 tot circa 1878 - kon Artis kaveltje voor kaveltje uitbreiden tot een oppervlakte van ruim 10 hectare. Honderdtwintig jaar lang moest de eerste dierentuin van Nederland het daar mee doen. Artis liet al in de jaren twintig van deze eeuw haar oog vallen op de toenmalige laad- en losplaats van de Nederlandse Spoorwegen aan de Plantage Doklaan, pal naast de dierentuin. Dit terrein had in die tijd een belangrijke functie van overslag van goederen. Via het aangrenzende water, het Entrepotdok, konden per spoor aangevoerde goederen met vrachtboten de grachten en/of het IJ bereiken.

Op de kaart van Pieter Mol uit 1772 ligt ten zuiden van de Plantage 'De Nieuwe Rapenburger Gragt'. Later werd deze gracht versmald door uitbreiding van het terrein dat nu 'Doklaan' heet: de gracht kreeg de naam 'Entrepotdok'. Met het Doklaanterrein zou Artis - echt Amsterdams - aan het water komen te liggen: een dierentuin aan de gracht.

Opeenvolgende Artis-directies wilden het terrein inlijven - het was immers de enige mogelijkheid voor de dierentuin om haar gebied verder te vergroten - maar om allerlei redenen mislukten die pogingen steeds.

Figure 4. Acceptable result, wrong method

Procedure

The experimental part of this study was spread out over three sessions for each class. During the first meeting the students were invited to take part in the pretest. Together with the ICT teacher or the home room teacher the experimenter introduced the study to the learners. They were also informed that they should experiment with the program before calling out for help. They were shown where on the computer they could find the map with their name and the Word documents they had to work with and how to save the data. Next they did the pretest for which they had 15 minutes. The first task of the pretest was an example and the experimenter guided the students through it to ensure that they could understand the exercise format. The experimenter advised them to follow the job-aid again and if they asked questions to solve the tasks they were ignored or advised to consult the job-aid and follow the instructions. Some examples of questions the students posed are “Did I do this procedure right?”, “Where was the function I need to use again?” and “My document is gone! What happened?”. The last question is concerned with the management of the study and the experimenter supported the students during these kind of problems.

The day after the pretest, the experimenter returned to the class for the training of 45 minutes and the immediate posttest of 15 minutes. The students each received a post-it with their name on it and a code which represented the condition they were in (e.g. “IM” for interactive version with meta cognitive assessment). The code was useful for checking whether the students worked with all the necessary material and were not confused by the actions of students from the other groups. The four groups were then addressed separately in order to ensure that they were using the correct version of the video, handed the correct job-aid and are clear on all the instructions.

The participants in the first condition, including the measurements on self-pacing and FOK, were instructed to think carefully about their own understanding and task performance as well as to make use of the self-pacing options if needed. The participants in the second and third condition, dealing with only measurements of self-pacing or FOK respectively, were shown how to fill in those questionnaires. In the control group, condition four, which had no additional assessment next to the task performance, the students were discouraged to try to pace the video tutorial.

After these instructions the students put on their headsets and started the training. Each task followed the pattern of: Read the exercise, Watch the video and then Practice it in the Word document.

When the students were finished with the training, or the 45 minutes were up, they were asked to close the video and continue on the computer with a fun or learning activity of their or their teacher’s choice. The experimenter checked if all the necessary measurements had been filled in. After about five minutes the children were handed out the posttest. The retention test was taken seven days after training and consisted of the same structure and instructions as the posttest. During the post- and retention-test the students were not allowed to access the videos. Both tests had a 15 minutes time limit in which all the learned procedures needed to be applied. At the end of the retention test the students were asked to give feedback about the whole experiment.

Analysis

The raw data was put into Microsoft Excel first for the ease of coding and a visual overview. Lists of the demographic data of the participants, which were collected from the teachers, the condition they were sorted into and the raw data of the experiment were then put into SPSS for statistical analysis. In order to analyze the effects of time and training on the procedural knowledge gain ANOVAs were computed during and on the two moments after training, using the pre-test scores as a covariate. Repeated Measures analysis was conducted with all the four test moments as within-subject factors and the motivational questionnaire data as a covariate to determine whether significant changes in motivation or cognition had taken place.

Levene‘s test of homogeneity analysis was included within all steps of the statistical analysis and the non-parametric Kruskal-Wallis Test and Jonckheere-Terpstra Test were computed if a check revealed a violation of the homogeneity assumption and to check for trends between groups. The degrees of freedom vary minimally on some variables, due to missing data. The analyses were all made two-tailed with an alpha of 0.05. For significant main results, pairwise comparisons are computed using Bonferroni's post-hoc test. Partial Eta-squared (η^2) was used to report effect size. The effect size tends to be regarded as small for $\eta^2 \approx 0.02$, medium for $\eta^2 \approx 0.13$ and large for $\eta^2 \approx 0.26$.

Results

Overview

Guiding the participants to assess their FOK states as well as giving them the opportunity to self-pace their video experience, were both expected to increase training effect. No significant difference was found between the conditions on any measurement during the entire experiment. The video tutorial significantly improved the participants' task success.

Cognitive outcomes and predictors

The Means and Standard Deviations shown within Table 1 indicate that participants started out with little task success. During training they performed exceedingly better with a mean of task success close to 75% in some conditions. The difference between the Pre-test and the Training success was statistically significant and substantial, $F(1,106) = 914.65$, $p < 0.001$, $\eta^2 = 0.896$. The effect size at this comparison is very large indicating that the improvement in task success did not only stem from the size of the sample group but from the video tutorial itself. It shows that the video tutorial was very successful as a job aid. A similar effect can be found when comparing Pre-test scores with the follow-up measurements. From Pre-test to Post-test there was a significant increase in task performance, $F(1,106) = 199.52$, $p < 0.001$, $\eta^2 = 0.653$. The same goes for the comparison between Pre-test to Retention-test, $F(1,106) = 240.83$, $p < 0.001$, $\eta^2 = 0.694$.

When looking at the difference between Training and Post-test it shows that task success decreased significantly from the former to the latter, $F(1,106) = 227.38$, $p < 0.001$, $\eta^2 = 0.682$. From Post-test to Retention-test the students' task success improved a little in some conditions, but significantly and with a small effect size, $F(1,106) = 6.42$, $p = 0.013$, $\eta^2 = 0.057$. Lastly, the difference in task success between the Training and the Retention-test scores proved to be significant as well, $F(1,106) = 123.36$, $p < 0.001$, $\eta^2 = 0.538$. Initial motivation and cognition showed to be no predictor for task success during training.

Table 1. *Means (and Standard Deviations) for Task Performance before, during and after Training*

	Pre-test	**Training**	**Post-test**	**Retention-test**
	Mean (SD)	*Mean (SD)*	*Mean (SD)*	*Mean (SD)*
1st Condition (SP + FOK)	17.2% (10.7)	73.6% (14.2)	47.1% (17.9)	57.1% (19.7)
2nd Condition (SP + no FOK)	19.1% (13.9)	70.1% (16.2)	43.9% (21.9)	45.5% (19.3)
3rd Condition (no SP + FOK)	19.8% (11.8)	73.2% (18.7)	47.6% (18.6)	49.7% (20.2)
4th Condition (no SP + no FOK)	16.8% (10.3)	72.8% (18.5)	44.5% (24.3)	51.1% (22.2)
Total	18.2% (11.7)	72.4% (16.8)	45.8% (20.6)	50.8% (20.5)

$n_{1st\ Condition} = 27$; $n_{2nd\ Condition} = 28$; $n_{3rd\ Condition} = 28$; $n_{4th\ Condition} = 27$; $n_{Total} = 110$;
SP = the Self-pacing functions within the video; FOK = the Feeling of Knowing assessment in the job-aid

The video tutorial proved to be an effective tool for the acquisition of procedural knowledge. Next to that, the students' performance on the Retention-test shows that the learning gain was lasting.

Table 1 further shows that there were no significant differences of Task success between conditions, $F(3,106) = 0.51$, $p = 0.680$, $\eta^2 = 0.014$. A repeated measures analysis was conducted on the four task performance measurements. The FOK assessment and the self-pacing options showed to have no effect on the cognitive outcomes. During training and in the follow-up tests the participants performance on the tasks did not depend on the condition, $F(3,106) = 0.26$, $p = 0.857$, $\eta^2 = 0.007$ for the Training, $F(3,106) = 0.21$, $p = 0.890$, $\eta^2 = 0.006$ for the Post-test and $F(3,106) = 1.51$, $p = 0.218$, $\eta^2 =$

0.041 for the Retention-test.

When the data was further explored by excluding one of the main variables (either FOK or Self-pacing) from the data set, the comparison of task performance between the two remaining conditions showed that there were no significant differences, $F(1,55) = 0.36$, $p = 0.549$, $\eta^2 = 0.007$ for Self-pacing and $F(1,54) = 0.00$, $p = 0.996$, $\eta^2 = 0.000$ for FOK. Prior knowledge also appeared to have no effect on the success of the Self-pacing options and the FOK assessment, $F(16,55) = 1.09$, $p = 0.397$, $\eta^2 = 0.315$ for Self-pacing and $F(15,54) = 1.13$, $p = 0.365$, $\eta^2 = 0.308$ for FOK.

Another important topic in this study is task complexity and it needed to be explored. The students struggled with some of the formatting procedures more than with others. Repeated measures analyses with the individual tasks of each of the four task performance measurements were conducted and indicated that differences in task complexity showed a similar trend on all four and across all conditions. Table 2 shows this trend within the Training tasks. The first few procedures, simple interaction with the “ruler” function, were easier to replicate than the final tasks, concerned with the automatically generated table of contents. The difference in task performance between the “easiest” and the “most difficult” task was significant, $F(1,105) = 165.08$, $p < 0.001$, $\eta^2 = 0.611$ during Training. Again the scores did not differ across conditions, $F(3,105) = 0.34$, $p = 0.797$, $\eta^2 = 0.010$.

When looking at all the tasks of all four cognitive measures individually it can be found that the only times the students’ scores did differ across condition was during the two Table of Contents-focused tasks during Training, $F(3,108) = 3.88$, $p = 0.011$, $\eta^2 = 0.100$ (for the Headings) and ,$F(3,108) = 2.93$, $p = 0.037$, $\eta^2 = 0.077$ (for the creation of the TOC), and the Create a List task during the Retention-test, $F(3,108) = 4.34$, $p = 0.006$, $\eta^2 = 0.110$. In the case of the Retention-test a slightly better task performance from the students of the first condition is shown on four out of the nine tasks (see graph in the appendix) which could mean that the combination of metacognitive monitoring and self-pacing functions during Training had some long-term effect on the procedural knowledge of these tasks, but the evidence is weak as no general effect of condition was found.

Table 2. *Means (and Standard Deviations) for Task Complexity during Training*

	1st Condition (SP + FOK)	**2nd Condition (SP + no FOK)**	**3rd Condition (no SP + FOK)**	**4th Condition (no SP + no FOK)**	**Total**
	Mean (SD)	*Mean (SD)*	*Mean (SD)*	*Mean (SD)*	*Mean (SD)*
Adjust Right Margin	96.3% (19.2)	96.5% (13.1)	88.9% (32.0)	98.2% (9.6)	95.0% (20.4)
Adjust Left Margin	83.4% (36.7)	76.8% (41.9)	90.8% (27.9)	77.8% (42.4)	82.1% (37.6)
Adjust Citation Right	84.6% (36.1)	78.1% (41.6)	88.9% (32.0)	87.2% (31.8)	84.6% (35.4)
Adjust Citation Left	97.9% (9.7)	92.7% (22.4)	92.6% (22.8)	91.9% (26.6)	93.8% (21.2)
Format Paragraphs	92.6% (22.8)	79.8% (39.1)	83.4% (34.0)	82.8% (36.5)	84.6% (33.6)
Adjust List	86.5% (32.7)	83.4% (33.3)	84.1% (33.2)	86.5% (29.6)	85.1% (31.8)
Use Tab for List	69.3% (46.2)	70.4% (45.7)	70.0% (43.9)	69.3% (44.0)	69.7% (44.3)
Select Headings	72.6% (44.4)	39.7% (43.9)	58.2% (45.3)	73.4% (34.3)	60.8% (43.9)
Create Table of Contents	52.5% (47.1)	82.8% (27.4)	71.7% (39.4)	61.2% (43.9)	67.2% (41.1)

$n_{\text{1st Condition}} = 27$; $n_{\text{2nd Condition}} = 28$; $n_{\text{3rd Condition}} = 27$; $n_{\text{4th Condition}} = 27$; $n_{\text{Total}} = 109$

To analyze which factors predict task performance a regression analysis was conducted with all cognitive measures and yielded the following results. The Pre-test scores fell short of predicting the students' performance during Training. The analysis did not result in a significant model, R^2 = 0.03, F(1,108) = 3.70, with the Pre-test scores not serving as a significant predictor (ß = 0.263, p = 0.057), but with a correlation of, r(1,109) = 0.18, p = 0.028, between the two variables. The Pre-test scores did however predict the Post-test performance, R^2 = 0.09, F(1,108) = 11.73, ß = 0.554, p < 0.001. Next to that the Pre-test scores did not predict the Retention-test scores, R^2 = 0.02, F(1,108) = 1.65, ß = 0.217, p = 0.201. It can thus be assumed that prior knowledge plays a role in knowledge acquisition and that it has greater effect on short-term learning than on the long-term. Training scores could not be used as a predictor for Post-test scores, R^2 = 0.29, F(1,108) = 44.33, ß = 0.663, p < 0.001, and Retention-test scores, R^2 = 0.18, F(1,108) = 23.04, ß = 0.513, p < 0.001. A regression analysis with Retention-test scores as dependent variable and the Post-test scores as the independent variable, R^2 = 0.23, F(1,108) = 32.50, ß = 0.480, p < 0.001, indicates that for a long-term learning effect the Training with the video and the Post-test had a positive connection and an impact on learning. Another regression analysis proved that condition did not predict task performance during Training, R^2 = 0.00, F(1,108) = 0.00, ß = 0.014, p = 0.963, and that there was no connection. The same goes for condition as a predictor of the other three cognitive measurements.

Table 3. *High, Middle and Low scoring percentages of students during the cognitive measures*

	Pre-test	**Training**	**Post-test**	**Retention-test**
High scoring	5.5%	67.9%	19.3%	23.9%
Middle scoring	31.0%	24.8%	39.4%	44.0%
Low scoring	61.5%	7.3%	41.3%	32.1%

High = more than 75 % task performance rate; *Middle* = between 50% and 75%; *Low* = less than 50 %; $n_{\text{Total}} = 109$

When checking for predictors for cognitive measurements with the High, Middle and Low scoring categorization visible in Table 3 it can be found that in all cases, except from Pre-test to Training, $R^2 = 0.02$, $F(1,108) = 1.62$, ß = 0.127, $p = 0.206$, could a former cognitive measurement be used as a predictor for a latter. Only five students could be categorized as High scoring during Pre-test, which is a fairly small part of the sample, but these continued to perform very well on the tasks of the latter cognitive measures, which could be a small indicator for the effectiveness of prior knowledge on learning. The video worked effectively on most students, which can be seen by the shift from Low to High scoring students from Pre-test to Training. The table also shows that in comparison to the Post-test the students performed slightly better on the Retention-test, which supports the idea that it is useful to have time reflect on one's learning.

Monitoring and Self-pacing outcomes and predictors

As the conditions did not differ significantly from each other on the cognitive outcomes, it is obvious that there is little need to describe the effect of the FOK effect and the effect of the interactive functions on learning. When it comes to FOK accuracy it was shown that no effect of FOK scores on task performance during the Post-test was found, $F(2,54) = 1.17$, $p = 0.320$, $\eta^2 = 0.044$, which indicated that the students could not accurately estimate their chances of replicating the procedures.

Another aspect, however, which should be regarded is the influence these variables have on each other. Correlational analyses shown that there were no significant connections between FOK and Self-pacing. When looking at FOK and the quantity in which the students made use of the interactive functions no connection was found, $r(1,27) = 0.19$, $p = 0.160$, and the same is true for the perceived effectiveness of the functions, $r(1,26) = 0.25$, $p = 0.110$.

While no general effect from the FOK assessment was found, analyses of variance revealed that there were some influences on individual tasks, $F(3,52) = 4.69$, $p = 0.006$, $\eta^2 = 0.227$ (for the Creation of Paragraphs and the accompanying FOK assessment) and $F(2,48) = 9.73$, $p < 0.001$, $\eta^2 = 0.297$ (for the Lists), but these were only encountered during Training and not on the follow-up tests.

A regression analysis from scores from the Self-pacing questionnaire on the users' quantity of use proved that they could be used as a predictor on the Post-test scores, $R^2 = 0.18$, $F(1,54) = 11.82$, ß = -2.799, $p = 0.001$. A correlational analysis from the same variables showed a negative correlation, $r(55,55) = -0.42$, $p = 0.001$, which indicates that the students who replayed the video tutorials less often scored higher on the Post-test and the other way around. This outcome is contrary to the results from the literature-study and cannot be considered valid as there were mistakes made during the construction of the Self-pacing questionnaire.

A positive correlation was found between the students' quantity of use of the Self-pacing options and the perceived effectiveness, $r(54,55) = 0.43$, $p = 0.001$. It can be assumed that the students who made more use of the functions valued them more.

Motivational outcomes and predictors

A one-way analysis of variance (ANOVA) was conducted on the three factors of the IEMQ and showed no significant differences between the conditions. On the Frequency subscale, $F(3,106) = 0.85$, $p = 0.470$, $\eta^2 = 0.023$, on the Relevance subscale, $F(3,106) = 0.75$, $p = 0.526$, $\eta^2 = 0.021$, and on the Self-efficacy subscale, $F(3,106) = 1.31$, $p = 0.274$, $\eta^2 = 0.036$.

Table 4 shows the means of the first two factors fairly close to the scale midpoint. The students' degree of Self-efficacy depicted a slightly higher than the midpoint confidence in their capacities to complete the Word formatting tasks.

Analyses of covariance (ANCOVA) on the students' Training scores, with the motivational factors as covariates, revealed no significant effect of condition.

Table 4. *Means (and Standard Deviations) for Motivational factors of the IEMQ during Pre-test*

	Frequency	**Relevance**	**Self-efficacy**
	Mean (SD)	*Mean (SD)*	*Mean (SD)*
1st Condition (SP + FOK)	3.23 (1.69)	3.36 (1.63)	4.21 (1.79)
2nd Condition (SP + no FOK)	3.56 (1.39)	3.55 (1.59)	4.49 (1.47)
3rd Condition (no SP + FOK)	3.16 (1.34)	3.02 (1.34)	3.91 (1.55)
4th Condition (no SP + no FOK)	2.95 (1.35)	3.12 (1.33)	3.73 (1.39)
Total	3.24 (1.45)	3.27 (1.48)	4.09 (1.57)

Scale values range from 1 to 7; the scale midpoint is 3.5; n= 110

Correlational and regression analyses of the three factors of the IEMQ as predictors of task performance on the Pre-test and Training scores (and the other two measures) yielded no significant results, except in the case of Self-efficacy and Pre-test scores which correlated positively, $r(1,109)$ =

0.20, p = 0.015. For Training Self-efficacy scores did not predict task performance, R^2 = 0.02, $F(1,108)$ = 2.06, ß = 0.295, p = 0.155.

The positive correlation between the scores from the Self-efficacy questions and the Pre-test scores shows that there is a connection between the degree learners can estimate their own skills and their actual skills.

Demographic outcomes

Looking at the influence of the demographic factors on the measurements taken in this study a few things appear noteworthy. Younger students show less confidence in their FOK states, $F(2,54) = 6.10$, $p = 0.004$, $\eta^2 = 0.193$. This could stem from their own idea that they are not as able as older students to know about their learning progress and have less confidence in their learning.

Furthermore, the results of a repeated measures analysis with the participants age as the independent variable indicate that there is a difference between students of different age and their task performance, $F(2,106) = 4.38$, $p = 0.015$, $\eta^2 = 0.076$.

When looking at each of the four test moments individually it shows that there are differences in task performance between the students of different ages on all four occasions, see Table 5, but only the ones during the Pre-test appear to be significant, $F(2,106) = 4.82$, $p = 0.010$, $\eta^2 = 0.083$. No effect of Gender was found on any measurement.

Table 5. *Means (and Standard Deviations) for Task Performance before, during and after Training*

	Pre-test	**Training**	**Post-test**	**Retention-test**
	Mean (SD)	*Mean (SD)*	*Mean (SD)*	*Mean (SD)*
10 Years	13.5% (9.0)	69.2% (18.6)	40.4% (21.9)	49.4% (20.2)
11 Years	18.9% (12.8)	72.2% (16.1)	44.9% (20.1)	47.3% (20.4)
12 Years	22.4% (11.0)	75.8% (14.9)	53.3% (18.7)	57.8% (20.2)
Total	18.2% (11.7)	72.4% (16.8)	45.8% (20.6)	50.8% (20.5)

$n_{10\ Years} = 32$; $n_{11\ Years} = 48$; $n_{12\ Years} = 29$; $n_{Total} = 109$

Discussion and Conclusion

The study looked at the general effectiveness of the video tutorials on the learner's task performance. The first hypothesis, the idea that the videos and accompanying job-aids have an influence on the efficiency of users' knowledge acquisition, was confirmed by the results. Before seeing the videos, during Pre-test, the students' task performance was at a low rate of 18 %, but rose to a high of 72 % during training. The video tutorials apparently instructed the students on how to deal with the tasks and they began to understand the formatting functions within *Word*. Whether the video helped the students recall procedural knowledge they had on the tasks before training or it did actually teach them something new is not necessarily important as the results indicate that the videos effectively supported task achievement either way. Motivational differences appeared not to affect the general support of the videos.

Although there was only a short amount of time (about five minutes) between the administration of the training and the Post-test, there was a 25 % decrease in task performance between the two mean scores. The reason for this rapid decrease could stem from the context in which the formatting tasks had to be handled. During training the participants watched a video tutorial and possessed the procedural knowledge to tackle the following task, while during the Post-test (and the Retention-test) they had to select the right strategy from all the video tutorials they had watched and fit it to the task. In training the students were guided through the different procedures one by one and did not need to focus on selecting the proper strategy. Another reason could be the amount of support. While instructions are given to the participant during training, the absence of support is important for the Post-test. If a participant has relied on the support too heavily or had problems integrating the new information into knowledge, it would be very difficult to select and perform the correct procedure during testing. The decrease in task performance can thus be accounted for by either the students' memory or the design of the materials.

On the two tests after training the mean task performance was moderately but significantly higher than during the Pre-test, with 46 % for the Post-test and 51 % for the Retention-test. Through this the results show that the students could perform the tasks without the immediate support of the video, although at a lower rate. The videos support the learning process effectively and have a continuing effect on learning.

The interesting fact that all four groups performed slightly better on the Retention-test than on the Post-test, was a surprise as no additional training was administered. The improvement could be explained by several factors. The students might have been a bit tired after the training and thus scored worse on the Post-test than if they would have had time to recuperate. The short break between training and Post-test was insufficient and may have reduced their motivation to perform well on the latter. It might have also been beneficial for the students to reflect on their newly acquired procedural knowledge for a week.

With an increase in task performance of more than 50 % from Pre-test to Training the instructional video has fulfilled its purpose, but the 50 % task success on the Retention-test also shows that there is still room for improvement. The design guidelines appear to have helped in the construction of the video and in its ultimate success, so what else could be done to improve the video even further?

When trying to answer this question it should be considered that the students in this study mainly came from low SES background which could have affected their learning behavior and diminished the effects of the video. Learners from middle or high SES background, like the ones confronted with the same video in a similar study (van der Meij and van der Meij, 2014), are expected to show higher task performance success (see Bradley and Corwyn, 2002). This study was the first to present this material to a sample group of students from low SES background and even though the effects of the video could have been higher the general effectiveness of the instructional video underlines its usefulness for all kinds of learners.

The research on task complexity showed that the tasks on the automatically generated table of contents were the most difficult for the students (see Table 2). They differed in their nature from the tasks on the “ruler” function. It might be advantageous to exclude them from the other tasks and put them into a separate video tutorial. Nembhard and Osothsilp (2002) showed in their study that higher task complexity significantly affects the variance of individual learning rates. It can thus be assumed that a training with tasks fitted to the learners’ general prior expertise will have the best learning effect. Results from the current study show that the Table of Contents-tasks were too complex for most students.

The role of prior knowledge also showed to have influence in this study, but the results for its effects are weak. The five students who scored high on the Pre-test and appeared to have some experience with *Word* also performed well on the other cognitive measures, but as there were only so few with that kind of experience the results of this analysis are not convincing. Results from the correlational analysis, on the other hand, prove that there was a positive correlation between scores from the Pre-test and Training which supports the importance of prior knowledge for learning. Pre-test scores also served as predictors for Training scores. Earlier studies showed the importance of prior knowledge on the learners’ learning gain and especially the behavior with new tasks (Dunbar and Klahr, 1988; Zimmermann, 2000). Although there are parts of the results that underline the role of prior knowledge in learning, we believe that the evidence would have been stronger with students with more computer experience.

The students appeared a bit tired and put out after each task performance measurement but the advice of Plaisant and Shneiderman (2005) to keep a video length of 15 to 60 seconds should be followed. We suspect that students wanted to get through the videos fast in order to perform well and be rid of their duties. In order to keep following the design guidelines in the construction of future videos and simultaneously decrease the students’ stress and task anxiety a computer training could be advised to increase general prior knowledge with computers which could in turn make it easier to deal with higher task complexity (see Chang, 2005).

Time on Task (TOT) was not measured and could have given insights in the students’ behavior with the procedures (see Son and Metcalfe, 2000). It could be assumed that the students with lower FOK ratings would have spent more time on the more difficult and time consuming tasks. The students in this study were given a time limit for their entire cognitive measurements, which means that they had no complete control over the time they could work on each task. Thiede and Dunlosky (1999) showed that FOK assessment can influence behavior as it shifts the focus between tasks. The most common literature finding on this effect is that students when assessing their FOK state spend more time on the difficult tasks than on the easy ones.

Furthermore the study investigated the effects added FOK assignment and self-pacing functions have on students' task performance as well as motivation during learning from a video tutorial. The second and third hypotheses, discussing the idea that students who were supported by both factors (encouraged to monitor their learning progress and option of pacing their learning from the video) would perform better than students with support from only one factor or no support, were not supported by the results of this study. The fact that no significant difference was found between any of the conditions that were applied during training on the measurements during and after the video tutorials, implies that the both factors did not have influence on the students' task performance.

This raises the question: Why was no effect of FOK assessment or Self-pacing on task performance found in this study when other studies show effects of them?

The relationship between the general effects of metacognitive thinking and learning discussed in CATLM could not be found in this study but is believed to exist. So how do we explain why the FOK assessment showed no effect? The answer is mostly centered on the students understanding of the FOK assessment. The children did not know how to engage in monitoring activities about their own learning and how helpful it can be for their progress. We suspect that they saw it as an exhausting exercise they had to get rid of, meaning that they did not effectively monitor their learning from the instructional video.

Next to that it needs to be considered that the effects of self-monitoring on procedural knowledge is not as well researched as on other types of knowledge and that no validated instrument exists to measure this relation (see Schneider, 2008). So did we use the right instrument? What would the right instrument for this material include? Although we assume that the type of metacognitive question and its frequency of appearance in the job-aid was an effective attempt to let the students monitor their knowledge we must ask: What would be alternative ways to measure the effect of FOK on learning? The instrument itself might not have been at fault but the whole way of measuring the FOK effect via textual instruments. An alternative and more effective way to measure the FOK effect might be the think-aloud method technique (see van Someren, et al., 1994; Zimmerman, 2008).

Schneider (2008) shows that procedural metacognitive knowledge differs clearly from conceptual and declarative metacognitive knowledge and advises strategy training procedures in different instructional domains. We speculate that the students in our study had no clear understanding on how they were supposed to assess their FOK on the learned procedures, even though they were instructed on it at the beginning of Training. They appeared not to be familiar with the technique and not aware of the benefits it can have for their task performance. A separate instructional session before the Pre-test in which the experimenter would model the technique and given little off-topic exercises could have made it clearer for the students. Although the video tutorials did model the desired behavior on the procedures to the students, we suspect that the modeling process was not "real" enough (not like a pedagogical agent, see van der Meij, van der Meij and Harmsen, 2012) in order to establish a connection to the effects of Bandura's motivational and learning theories. We also could not find a real connection between Bandura's theories and FOK states made by other researchers.

Furthermore, there are no studies with special FOK assessment training, which would model to the learners how they successfully assess their own FOK states during learning in a SDL environment. We assume that it would be efficient to include a half hour of FOK assessment training before the training with the software takes place. We advise to include this concept in future FOK studies on procedural learning.

Next to the missing experience with skills concerned with cognitive monitoring the students' inefficient interaction with their FOK states could stem from the cognitive exhaustion and the feeling that they have to do extra work and they thus associate it with stress and motivation dwindles, meaning that future studies should also control for stress which can influence metacognitive monitoring (Schwarz, 2010).

According to Perrotin, Tournelle and Isingrini (2008) the major part of variance in FOK accuracy is accounted by the learner's ability of the executive functioning cognitive processes, and especially the monitoring process of *shifting*, which describes the learner's individual ability to move back and forth between multiple tasks, operations, or mental sets (Monsell, 1996). No connection between the students' perceived FOK states and their follow-up scores was found in this study, which could be interpreted as the students' inability to monitor their learning, but we expect this to result from methodological constraints within the set-up of the study. We expect that there is variance in FOK accuracy between individuals, but it can only be explored if the learners know how to monitor their learning by themselves. It appears that the students in the current study were not familiar enough with the concept. It would have been important to give the students a deeper understanding of "why" they had to select a particular procedure to tackle a particular task. The rules for selecting the correct formatting procedure would have helped the students in future endeavors, but could not be included in this training as it would have exhausted the participants and blown the time constraints of school and experiment. To gather better information on metacognitive processes and their effects on learning and behavior with the tasks think aloud protocols could reveal deeper insights into thought processes. Self-reporting and specific questionnaires should also be considered in future studies on monitoring, but would have put too much cognitive strain (overload) on the participants of this study.

The workings of the Good Information Processing Model (Pressley, et al., 1989), mentioned in the theoretical background of this study, could thus not be replicated, due to the students' lack of understanding of the FOK effect.

The literature on Self-pacing in instructional videos (see Mayer and Chandler, 2001; Tversky, et al., 2002; Mayer, 2005; Moreno, 2006; and Ertelt, 2007) supports its importance for the success of videos. As individuals differ from another in what they can absorb and retain during learning it is important to give them control over their pace.

The eight guidelines (van der Meij and van der Meij, 2013) for the construction of interactive video tutorials proved to be useful as Training had a strong effect on the learners. The Self-pacing functions make part of these guidelines. The reason why the interactive components appeared to have no effect within this study comes from the fact that the technical equipment of the schools was not optimal for this kind of study, which led to the development of a Self-pacing questionnaire with several shortcomings. First, the questionnaire measuring the quantity of the use of the Self-pacing functions was too short and used a far too broad scale, resulting in scores that do not reflect the actual use of the video. Second, there was no understanding of "why" the interactive functions are useful for learning. And deeper questionnaire focus on the Self-pacing options together with more information on the workings for the learners could replicate the results of earlier studies. It is also questionable whether the self-pacing options presented in the job-aid are presented obvious enough for the children influenced by YouTube and the "click and lean back"-thinking.

Furthermore, there is a difference between being conscious of an option and actually making use of it. In order to see if the students really make use of the Self-pacing options future studies should be equipped with new computers to support screen capture programs or log files. These screen capture technologies could also tell the experimenter not only the behavior but also the time each student allocates to each task (TOT), which was also discussed above, which could support or disconfirm the findings of Son and Metcalfe (2000) on the way FOK assessment affects behavior, but this is a very time-consuming task for data collection and a log-file would be simpler. Pongnumkul and colleagues (2011) show how screen capture technologies can be incorporated into an SDL environment with instructional video.

We suspect that the contradictory negative correlation found in this study between the Self-pacing scores and the Post-test scores does not reflect a real connection but rather stems from problems with the self-made Self-pacing questionnaire and from misunderstandings of the job-aid. It was expected that the job-aid would guide the students' behavior in a certain way, but it cannot be assured that the learners behaved that way. After seeing a segment of the video tutorial the participants of the first condition (SP + FOK) read a short paragraph which explained the Self-pacing options and the fact that could make use of them "at all times". Then followed the FOK assessment and ultimately the work on the task within *Word*. The order in which these parts of the job-aid were put in might not have been optimal. The "Replay possible" paragraph should be repeated after the FOK assessment. We assume that the FOK assessment would influence the students' behavior (if they felt uncertain, they would replay parts of the tutorial), but after they answered the metacognitive question they might have already forgotten about the interactive options. For future studies we advise to make the possibilities clear through training on the concepts or a job-aid that was tested beforehand.

On a theoretical basis one might find a connection between the two concepts when looking at the apprehension principle from Tversky and colleagues (2002), stating that animations should be readily and accurately perceived and comprehended together with Lang's Limited Capacity Model (2000) of mediated message processing, focused on the way the flow of video information challenges the user to decide which information to encode, process and store, from the background of interactivity and the monitoring process of *shifting* (described above) from the background of FOK effects. The Self-pacing options give the learner more possibilities to move back and forth between the sub-tasks of the procedures which might thus support *shifting*, which in turn might increase FOK accuracy.

The question remains: How can this type of learning from video be adjusted to fit to more authentic learning scenarios? It is sure that it would be advantageous to establish SDL environments over a longer time period to be effective and give the students a chance to get to know the strategies of a domain and that the best ways to measure the students' behavior in such an environment are the most unobtrusive ones.

The results from the IEMQ indicate that the students were neither highly nor under motivated by the tasks of the Pre-test. Their motivational scores are all fairly close to the scale midpoint. Although the students displayed slightly higher than average self-efficacy in the IEMQ they asked lots of questions during the Pre-test and showed that they had no idea how to perform the tasks they were asked to do. As described in the Procedure section the student's questions were not answered contextually and no solution was given, in particularly lost cases they were send on the right path through the job-aid. Their confusion was reflected in the task performance scores.

The positive correlation between the scores from the Self-efficacy questions and the Pre-test scores lets us believe that the learners in this study had an idea of what their actual skills were before Training. The analysis of the IEMQ results does not give many points worth discussing as it was only included in the Pre-test job aid, which was a methodological shortcoming. The IEMQ should have been administered not only during Pre-test, but also during Training in order to get a comparison. We now do not know how the students felt about the video tutorials. It might have also been useful to measure motivation during Post and Retention-test to look for long-term effects. This study could not research which of the eight design guidelines affected the user's motivation. No information was collected on how the students' motivation influenced their decision to engage in task execution and how this contributes to learning.

Results from the analysis of the influence of demographic factors show that age differences and their effects need to be discussed. Younger students show lower confidence in their acquired knowledge via monitoring during training. The younger students might not have been as experienced in the use of *Word* or computers in general as their older classmates. Prior (domain-specific procedural) knowledge was certainly a beneficial factor in the older students. The task performance measure from the Pre-test supports that idea. While Butterfield, Nelson, and Peck (1988) indicated that even preschoolers could make accurate FOK judgments, Schneider and Lockl (2002) showed that FOK rates demonstrate a small developmental progression. In a later study Schneider (2008) proved that metacognitive processes of self-monitoring improve as the child matures, but no clear differentiation has been given on what can be expected from a pupil at a given level. This study showed that younger students are less certain of their monitoring states and it might have found bigger differences between conditions when conducted with older (13 to 15 years) participants.

Some of the unexpected results, like the fact that there was no effect of condition, could also partly be explained by factors of the learners' social environment, concerned with the way the student's behavior is influenced by their peers. We suspect that students have a habit of committing to knowledge they do not have in order to not appear "stupid" in front of their classmates and some appeared to take the experiment as a simple quiz in which they wanted to be the best and tried to finish extra fast. These forms of behavior can manipulate any real score. It is of importance to keep in mind that the research has been conducted in three primary schools of a suburb with low SES, so the same study might have yielded different results in an area with higher SES, as it has been proven that SES is tied to learning and cognitive development (Bradley and Corwyn, 2002). In the case of the three schools the general computer training advised before could have a great impact. The area's SES has also influenced the methodology of the study as unfortunately no student logs (or screen captures) could be collected, as the demanding software would have crashed the old school computers. Some students already had problems with much more essential points as being able to listen to the recordings in good quality due to low-quality headphones.

The present study contributes to existing literature about instructional design and multimedia learning by supporting the success well designed instructional videos can have on learning procedures. And it does influence learning in schools as it was shown again that video tutorials are a valuable tool in instruction in a SDL environment. This answered only the first research question, but advice was given on how to conduct similar studies to answer the remaining questions.

In summary, it can be concluded that although this study did not support the idea that FOK and SP-options support learning from instructional video, we see that methodological constraints hindered the effectiveness of this study and that the literature supports the importance of these concepts for learning meaning that more research is needed.

In any case, two questions remain: What are the future problems to be tackled when it comes to learning from video? How can metacognitive thinking be introduced into the everyday lessons of the classroom environments and also made a SDL strategy the learners apply by themselves?

No harm has come to the participants during this study and no debriefing was necessary as the study's mechanisms were explained to them during the first introduction.

References

Azevedo, R., & Cromley, J. G. (2004). Does training on self-regulated learning facilitate students' learning with hypermedia?. *Journal of educational psychology*, *96*(3), 523.

Biggs, J., & Tang, C. (2007). Teaching for quality learning at university (Society for Research into Higher Education).

Blake, M. (1973). Prediction of recognition when recall fails: Exploring the feeling-of-knowing phenomenon. *Journal of Verbal Learning and Verbal Behavior*, *12*(3), 311-319.

Blumberg, P. (2000). Evaluating the evidence that problem-based learners are self-directed learners: A review of the literature. In D. H. Evensen & C. E. Hmelo (Eds.), Problem-based learning: A research perspective on learning interactions (pp. 199-226). New Jersey: Lawrence Erlbaum Associates, Publishers.

Bolhuis, S. (1996). Towards Active and Selfdirected Learning. Preparing for Lifelong Learning, with Reference to Dutch Secondary Education. Paper presented at the Annual Meeting of the American Educational Research Association (New York, NY, April 8-12, 1996).

Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual review of psychology*, *53*(1), 371-399.

Brown, A. S. (1991). A review of the tip-of-the-tongue experience. *Psychological bulletin*, *109*(2), 204.

Butterfield, E. C., Nelson, T. O., & Peck, V. (1988). Developmental aspects of the feeling of knowing. *Developmental Psychology*, *24*(5), 654.

Candy, P. C. (1991). *Self-Direction for Lifelong Learning. A Comprehensive Guide to Theory and Practice*. Jossey-Bass, 350 Sansome Street, San Francisco, CA 94104-1310.

Chang, S. E. (2005). Computer anxiety and perception of task complexity in learning programming-related skills. *Computers in Human Behavior*, *21*(5), 713-728.

Cennamo, K. S. (1994). “Sprouting” the Couch Potato: Techniques to Increase the Effort Learners Invest in Video-Based Materials. *Performance Improvement Quarterly*, *7*(2), 62-80.

Corno, L. (1992). Encouraging Students to Take Responsibility for Learning and Performance. *Elementary School Journal*, 93(1), 69-83.

Dunbar, K., & Klahr, D. (1989). Developmental differences in scientific discovery processes. *Complex information processing: The impact of Herbert A. Simon*, 109-143.

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* (Doctoral dissertation, Universitätsbibliothek Freiburg).

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. American Psychologist, 34, 906-911.

Fischer, G., & Scharff, E. (1998). Learning technologies in support of self-directed learning. *Journal of Interactive Media in Education*, *1998*(2).

Garrison, D.R. (1997). Self-directed learning: Toward a comprehensive model. *In Adult Education Quarterly*, 48(1) 18.

Harp, S. F., & Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology*, *90*(3), 414.

Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of educational psychology*, *56*(4), 208.

Hart, J. T. (1967), Memory and the memory-monitoring process. *Journal of Verbal Learning and Verbal Behavior*, 6, 685-691.

Hiemstra, R. (1994). Self-directed learning. *The sourcebook for self-directed learning*, 9-20.

Kennedy, G., Kraus, K., Gray, K., Judd, T., Bennett, S., Maton, K., Dalgarno, B. and Bishop, A. (2006), “Questioning the next generation: a collaborative project in Australian higher education,” paper presented at the 23rd Annual Ascilite Conference “Who’s Learning? Whose Technology?”, University of Sydney, Sydney.

Knowles, M. S. (1975). *Self-directed learning*. New York: association Press.

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological review*, *100*(4), 609.

Koriat, A. (1995). Dissociating knowing and the feeling of knowing: Further evidence for the accessibility model. *Journal of Experimental Psychology: General*, *124*(3), 311.

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological review*, *103*(3), 490.

Koriat, A., Ma'ayan, H., & Levy-Sadot, R. (2001).The intricacies of monitoring and control processes: The dual function of study time in self-paced learning.

Lang, A. (2000). The limited capacity model of mediated message processing. *Journal of communication*, *50*(1), 46-70.

Luke, B., & Hogarth, K. (2011). Developing and enhancing independent learning skills: Using video tutorials as a means of helping students help themselves. *Accounting Research Journal*, *24*(3), 290-310.

Martinez, M. E. (2006). What is metacognition?. *Phi delta kappan*, *87*(9), 696.

Mayer, R. E. (2005). *The Cambridge handbook of multimedia learning*. New York, NY: Cambridge University Press.

Mayer, R. E., & Chandler, P. (2001). When learning is just a click away: Does simple user interaction foster deeper understanding of multimedia messages?. *Journal of educational psychology*, *93*(2), 390.

Mayer, R. E., Dow, G. T., & Mayer, S. (2003). Multimedia learning in an interactive self-explaining environment: What works in the design of agent-based microworlds?. *Journal of Educational Psychology*, *95*(4), 806.

McCombs, B., & Vakili, D. (2005). A learner-centered framework for e-learning. *The Teachers College Record*, *107*(8), 1582-1600.

Monsell, S. (1996). Control of mental processes. *Unsolved mysteries of the mind: Tutorial essays in cognition*, 93-148.

Moore, M. G. (1973). Toward a theory of independent learning and teaching. *The Journal of Higher Education*, 661-679.

Moreno, R. (2006). Does the modality principle hold for different media? A test of the method-affects-learning hypothesis. *Journal of Computer Assisted Learning, 22*, 149-158.

Moreno, R. (2009). Learning from animated classroom exemplars: The case for guiding student teacher's observations with metacognitive prompts. *Educational Research and Evaluation, 15*(5), 487-501.

Moreno, R., & Mayer, R. (2007). Interactive multimodal learning environments. *Educational Psychology Review*, *19*(3), 309-326.

Morrow, L. M., Sharkey, E., & Firestone, W. A. (1993). *Promoting independent reading and writing through self-directed literacy activities in a collaborative setting*. National Reading Research Center.

Nelson, T. O. (1990). Metamemory: A theoretical framework and new findings. *Psychology of learning and motivation*, *26*, 125-173.

Nelson, T. O. (1996). Consciousness and metacognition. *American psychologist*, *51*(2), 102.

Nembhard, D. A., & Osothsilp, N. (2002). Task complexity effects on between-individual learning/forgetting variability. *International Journal of Industrial Ergonomics*, *29*(5), 297-306.

Perrotin, A., Tournelle, L., & Isingrini, M. (2008). Executive functioning and memory as potential mediators of the episodic feeling-of-knowing accuracy. *Brain and cognition*, *67*(1), 76-87.

Peterson, P. L., Swing, S. R., Braverman, M. T., & Buss, R. R. (1982). Students' aptitudes and their reports of cognitive processes during direct instruction. *Journal of Educational Psychology*, *74*(4), 535.

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

Pongnumkul, S., Dontcheva, M., Li, W., Wang, J., Bourdev, L., Avidan, S., & Cohen, M. F. (2011, October). Pause-and-play: automatically linking screencast video tutorials with applications. In *Proceedings of the 24th annual ACM symposium on User interface software and technology* (pp. 135-144). ACM.

Pressley, M., Borkwski, J. G., & Schneider, W. (1989). Good information processing: What it is and how education can promote it. *International Journal of Educational Research*, *13*(8), 857-867.

Race, P. (2005). Making learning happen. *A Guide for Post-compulsory Education*.

Reder, L. M., & Ritter, F. E. (1992). What determines initial feeling of knowing? Familiarity with question terms, not with the answer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(3), 435.

Schneider, W. (2008). The development of metacognitive knowledge in children and adolescents: Major trends and implications for education. Mind, Brain, and Education, 2, 114-121

Schneider, W., & Lockl, K. (2002). 10 The development of metacognitive knowledge in children and adolescents. *Applied metacognition*, 224.

Schneider, W., & Lockl, K. (2008). Procedural metacognition in children: Evidence for developmental trends. *Handbook of metamemory and memory*, *14*, 391-409.

Schraw, G. (1995). Measures of feeling-of-knowing accuracy: A new look at an old problem. *Applied Cognitive Psychology*, *9*(4), 321-332.

Schwartz, B. L. (1994). Sources of information in metamemory: Judgments of learning and feelings of knowing. *Psychonomic Bulletin & Review*, *1*(3), 357-375.

Schwartz, B. L., & Metcalfe, J. (1992). Cue familiarity but not target retrievability enhances feeling-of-knowing judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(5), 1074.

Schwarz, N. (2010). Meaning in context: Metacognitive experiences. *The mind in context*, 105-125.

Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(1), 204.

Swartz, R. J., & Perkins, D. N. (1990). *Teaching Thinking: Issues and Approaches. Revised Edition. The Practitioners' Guide to Teaching Thinking Series*. Midwest Publications, Critical Thinking Press and Software, PO Box 448, Pacific Grove, CA 93950-0448; tele.

Temple, C., & Rodero, M. L. (1995). Active Learning in a Democratic Classroom: The" Pedagogical Invariants" of Celestin Freinet (Reading around the World). *Reading Teacher*, *49*(2), 164-67.

Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of experimental psychology: Learning, Memory, and Cognition*, *25*(4), 1024.

Thiede, K. W., Anderson, M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of educational psychology*, *95*(1), 66.

Tversky, B., Morrison, J. B., & Betrancourt, M. (2002). Animation: can it facilitate?. *International journal of human-computer studies*, *57*(4), 247-262.

Van der Meij, H., Meij, J., & Harmsen, R. (2012). Animated pedagogical agents: do they advance student motivation and learning in an inquiry learning environment?.

Van der Meij, H. (2013). Motivating agents in software tutorials. *Computers in Human Behavior, 29*, 845-857.

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

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

Van Someren, M., Barnard, Y.J., Sandberg, J.A.C. (1994). *The Think Aloud Method: A Practical Guide to Modelling Cognitive Processes,* Academic Press, London.

Vermunt, J. D., & Vermetten, Y. J. (2004). Patterns in student learning: Relationships between learning strategies, conceptions of learning, and learning orientations. *Educational Psychology Review*, *16*(4), 359-384.

Vollmeyer, R., & Rheinberg, F. (2006). Motivational effects on self-regulated learning with different tasks. *Educational Psychology Review*, *18*(3), 239-253.

Wright, D. B. (1996). Measuring feeling of knowing: Comment on Schraw (1995). *Applied Cognitive Psychology*, *10*(3), 261-268.

Zheng, R., McAlack, M., Wilmes, B., Kohler-Evans, P., & Williamson, J. (2009). Effects of multimedia on cognitive load, self-efficacy, and multiple rule-based problem solving. *British Journal of Educational Technology*, *40*(5), 790-803.

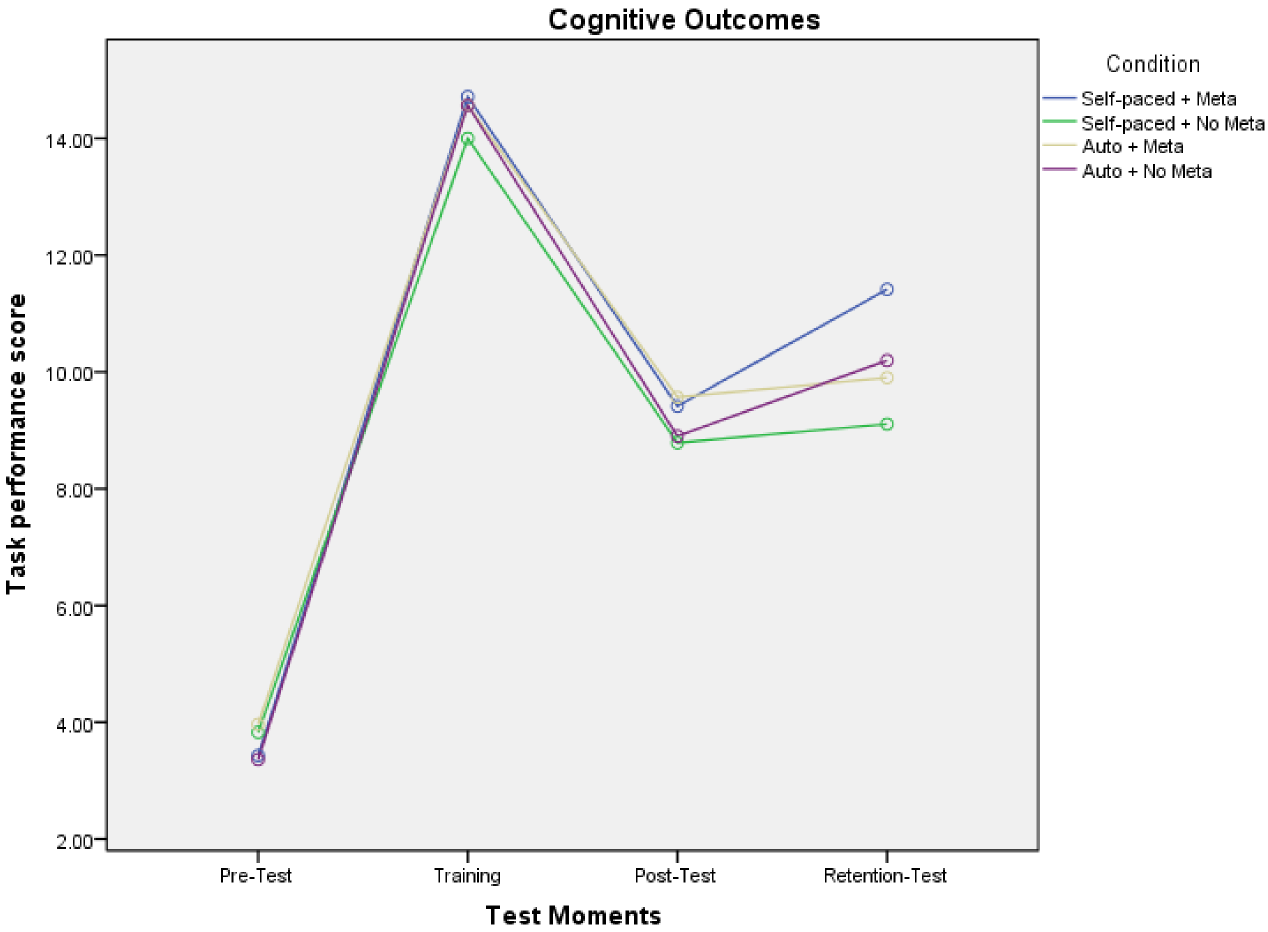
Zimmerman, B. J., & Martinez-Pons, M. (1990). Student differences in self-regulated learning: Relating grade, sex, and giftedness to self-efficacy and strategy use. *Journal of educational Psychology*, *82*(1), 51.

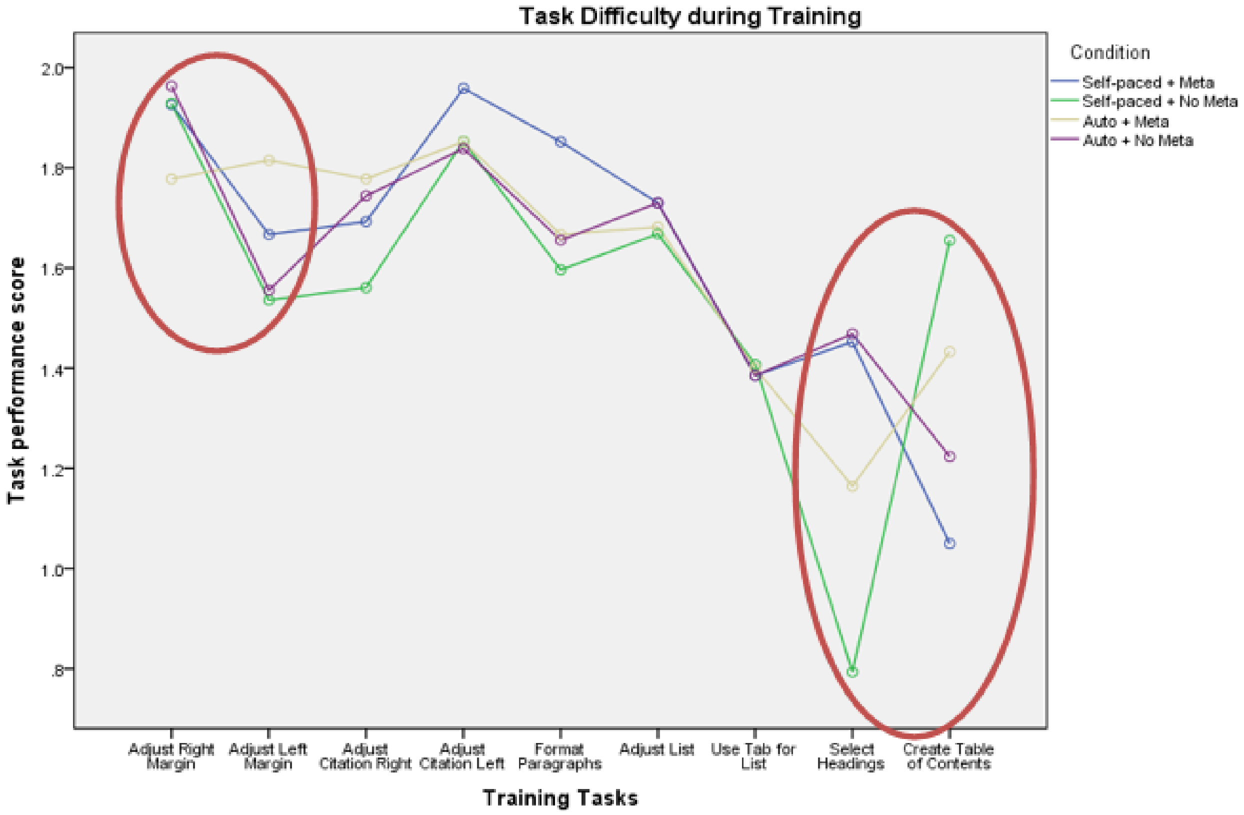
Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal*, *45*(1), 166-183.

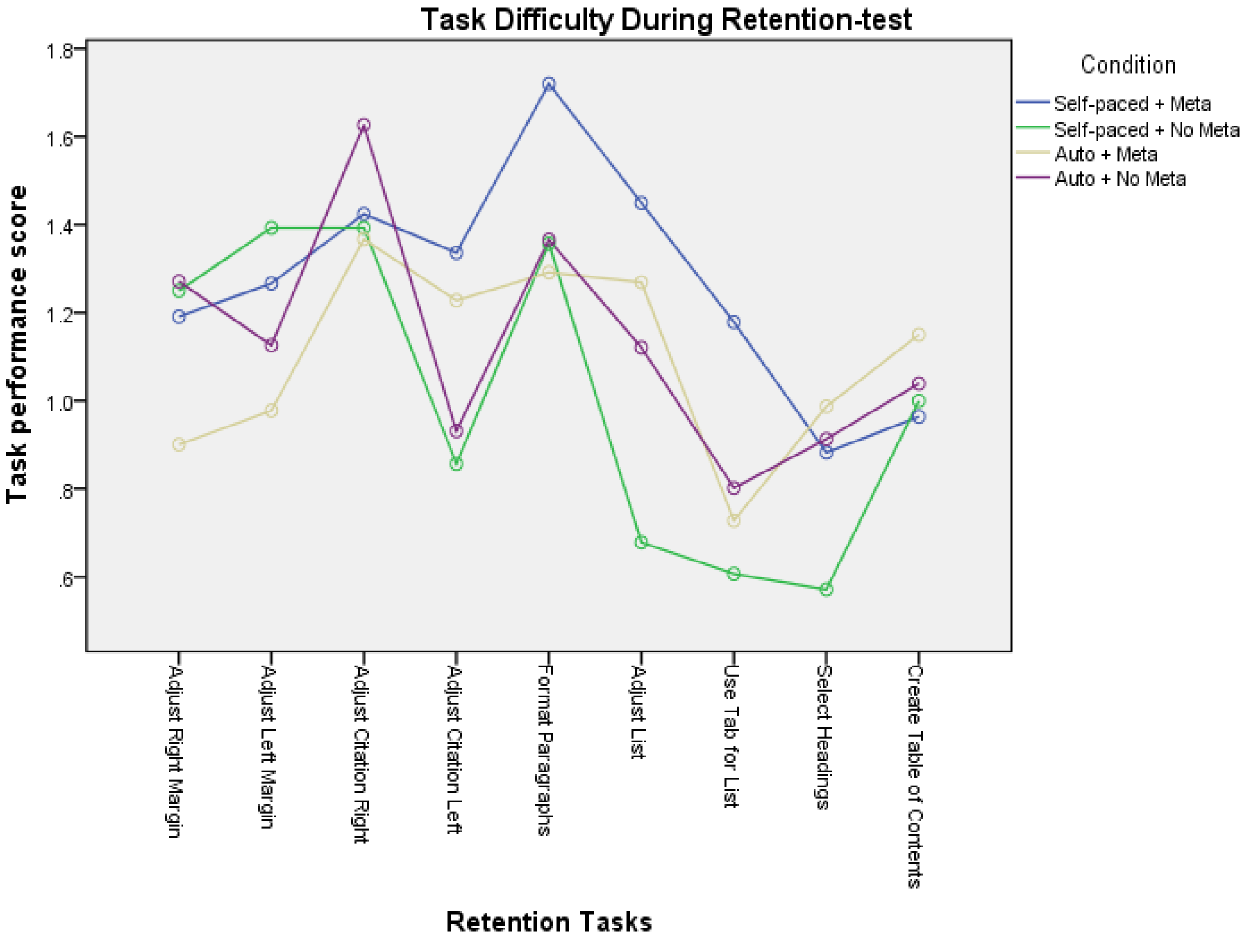
Zimmerman, C. (2000). The development of scientific reasoning skills. *Developmental Review*, *20*(1), 99-149.

Appendix

The graphs




Task Difficulty during Training
Condition
Self-paced + Meta
Self-paced + No Meta
Auto + Meta
Auto + No Meta
Task performance score
2.0
1.8
1.6
1.4
1.2
1.0
.8
Adjust Right Margin
Adjust Left Margin
Adjust Citation Right
Adjust Citation Left
Format Paragraphs
Adjust List
Use Tab for List
Select Headings
Create Table of Contents
Training Tasks


Task Difficulty During Retention-test
Condition
Self-paced + Meta
Self-paced + No Meta
Auto + Meta
Auto + No Meta
Task performance score
1.8
1.6
1.4
1.2
1.0
.8
.6
Adjust Right Margin
Adjust Left Margin
Adjust Citation Right
Adjust Citation Left
Format Paragraphs
Adjust List
Use Tab for List
Select Headings
Create Table of Contents
Retention Tasks