

Thesis

The effects of enriching instructional videos with retrieval practice on motivation and learning

Kim van Lieshout

6406416

07-06-2021

First assessor: Vincent Hoogerheide

Second assessor: Gesa van den Broek

Word count: 7413

Abstract

Retrieval practice is an effective learning strategy in which students have to actively retrieve information from their memory which improves long-term retention. The goal of this study was to investigate if the enrichment of instructional videos with retrieval practice increases the motivation and learning of primary school students, because the effects of retrieval practice on motivation and children is small. Additionally, this study focus on the unique relation of retrieval practice in video-based learning. In total, 95 children from the fifth and sixth grade have studied an instructional video. One group engaged in retrieval practice in-between watching the instructional video, one group engaged in retrieval practice after watching the video, and one group only watched the video three times. After the learning phase, participants completed an immediate and delayed posttest. This study showed that retrieval practice does not influence performance on an immediate posttest, which is in line with previous research. Surprisingly, the results shows no significant effect on performance on the delayed posttest. Besides, participants in the retrieval practice conditions did not have a higher situational interest, which was measured before and after the learning phase, although it is argued that retrieval practice is more engaging than restudying.

Introduction

Instructional videos are considered one of the most popular ways to deliver instruction (De Koning et al., 2018; Leisner et al., 2020). These videos provide students with information about certain concepts or phenomena. Instructional videos can be used in different learning environments, such as traditional courses, blended courses and massive online open courses (De Koning et al., 2018). Due to the current global COVID-19 pandemic, the use of instructional videos is now even more important since universities teach at distance. However, there is conflicting evidence about the learning outcomes of video-based learning compared to text-based learning (Leiser et al., 2020). Watching instructional videos once or multiple times can be seen as a passive cognitive activity. During a passive cognitive activity, students watch a video without engaging in any overt behaviour related to learning (Chi, 2009). Students find it difficult to concentrate while engaging in a passive activity, because they are easily distracted and tend to mind-wander (Szpunar et al., 2013).

Retrieval practice can be used in order to make studying instructional videos less passive, because students need to actively reconstruct knowledge (Karpicke, 2012). Retrieval practice is the act of calling information to mind instead of rehearing or rereading this information (Roediger & Butler, 2011). In a recent study of van der Zanden (2020), students in the fifth and sixth grade of primary school studied instructional videos at home in an online environment. The students in the retrieval practice conditions had to engage in retrieval practice by answering a free recall question. These students had to write down all the content and concepts they could remember. Students scored higher on the factual knowledge questions (i.e., retention) of the direct posttest when they engaged in retrieval practice in-between watching the instructional videos compared to students who did not engage in retrieval practice or engaged in retrieval practice after watching the instructional videos. However, no significant effect was found for the conceptual knowledge questions (i.e.,

comprehension). Although van der Zanden (2020) predicted that differences in performance on the direct posttest could be caused by mind-wandering, no effect was found among the conditions on mind-wandering.

The present study will focus on the relationship between instructional videos and retrieval practice by replicating and extending the study of van der Zanden (2020). This study is the first study which suggests positive results of retrieval practice in a video-based learning environment on learning with primary school children. As replication is important for the credibility of science, this study needs to be replicated. Since there has been a replication crisis in the social sciences (Earp & Trafimow, 2015), this study aims to contribute to this gap in replication studies. Moreover, children love to watch instructional videos (Beheshti et al., 2018), but it is important to know how to make learning from videos the most efficient. This study extends the study of van der Zanden (2020) with a delayed posttest in order to investigate the long-term effects on retention and comprehension of the learning materials. Additionally, this study will include an investigation about the motivation of the students. Moreover, the time-on-task between the conditions will be equal in order to investigate if the significant positive effect remains when students in the control condition spend the same amount of time on the task as students in the retrieval practice conditions. Before a detailed explanation of this study is given, the effects of retrieval practice on video learning and student motivation will be explained.

Theoretical framework

What is retrieval practice?

Retrieval practice is an instructional strategy in which individuals actively retrieve and reconstruct information from their long-term memory (e.g., quizzes or free recalls) without rehearing or rereading this information (Roediger & Karpicke, 2006). With free recall, students have to write down or say out-loud all the information they remembered from the

learning materials (Roediger & Karkpicke, 2006; Rowland, 2014). This retrieval leads to elaboration of their memory trace in which students can connect the new information to their prior knowledge (Roediger & Butler, 2011). When students reactivate their knowledge from their long-term memory, it is more likely they will be able to successfully retrieve the information in the future (Roediger & Butler, 2011). The retrieval practice effect is also known as the testing effect (Moreira et al., 2019; Roediger & Butler, 2011). According to the ICAP framework of Chi and Wylie (2014), retrieval practice leads to more learning than restudying. In this framework, the overt behaviour of students can predict their learning outcomes. The learning of students increases when they are more engaged with the learning materials, from passive to active to constructive to interactive. Restudying a video is a passive activity, because students just watch the video without doing anything else. However, retrieval practice is an active activity, because students have to put in effort during a retrieval practice activity, which leads to higher learning outcomes (Chi & Wylie, 2014; Roediger & Butler, 2011).

There is a substantial agreement between researchers about the beneficial effects of retrieval practice for long-term memory, whilst the effects of retrieval practice immediately after the learning phase are not noticeable (Abel & Bäuml, 2020; Roediger & Butler, 2011). However, there is some debate about whether retrieval practice leads to short-term effects. A study of van den Broek et al. (2014) shows that the testing effects on the recall of word pairs effected long-term memory. Moreover, when the participants indicated they successfully retrieved the word pairs, the testing effect was already significant after practice (van den Broek et al., 2014). Thus, this study indicates that the effects of retrieval practice can already be immediately noticeable after the learning phase. According to Roediger & Butler (2011), most research shows that retrieval practice improves retention, but retrieval practice also enhances the comprehension of materials, although this effect is less strong. Besides, retrieval

practice can provide insight in the learning process (Karpicke, 2012). When students engage in retrieval practice, they will be provided with metacognitive knowledge. This knowledge gives learners the opportunity to focus on information which they do not master yet, which improves their self-regulated learning (Szpunar et al., 2014).

Retrieval practice and instructional video learning

Above shows that the effect of retrieval practice on learning is positive. Yet, the amount of research about retrieval practice in video-based learning is not extensive (Cummins et al., 2016). Besides the study of van der Zanden (2020), there are some other studies which found promising evidence about the effectiveness of retrieval practice in video-based learning. A study of van der Zee et al (2018), focused on the performance of students in a massive online open course when retrieval practice was integrated in the materials. The participants had to read a summary or write their own summary after they watched an instructional video without access to the learning materials. They found a positive effect on performance on the final test when students read a summary or wrote their own summary than students who did not engage in any activity after watching a video. The control group of this study spent less time studying the materials, which could also bias the positive effect.

Moreover, Szpunar and colleagues (2013) found positive evidence for the integration of retrieval practice in online lectures. Using 80 university students, the study showed that participants who answered test questions after each lecture segment, experienced less mind-wandering and had a higher performance than participants who only restudied the lectures. More specific, the participants who had to answer test questions were also encourage to take notes. Taken this into account, the effect of only answering test questions on the performance remains unclear. Yong and Lim (2016) also found positive evidence for test questions integrated with videos. In this study, participants who studied the video-recorded lecture repeatedly (i.e., three times) performed better on the immediate posttest than participants who

tested themselves repeatedly (i.e., once watching the video and taking three tests) after the video lecture. In contrast, this effect changed over time. Participants who tested themselves repeatedly performed better on the delayed posttest than participants who studied the video-recorded lecture three times (Yong & Lim, 2016).

In short, there are some promising studies regarding the effects of retrieval practice in video-based learning. However, there are some questions regarding this effect. First, these the studies conducted their research with college students. According to Goossens et al. (2013) and Moreira et al. (2019) the effects of retrieval practice on learning with primary school children remains unclear. The development of children during primary school is critical. During these ages, children are expected to begin to implement learning strategies on their own, facing many difficulties by executing effective strategies (Pressley & Harris, 2009). Because of this, the learning strategies of children are not comparable to college students yet, which can influence the effect of retrieval practice on their learning. Second, these studies do not provide evidence when to offer retrieval practice during the learning phase. However, retrieval practice in-between the study phase could help students to become aware of their potential knowledge gap in an early stage. They can focus on these knowledge gaps during the rest of the study phase, which could improve their performance. Only one study showed that these learning activities (i.e., retrieval practice or explaining study content) after the study phase are less effective than learning activities in-between the study phase (Lachner et al., 2019).

Effects of retrieval practice on student motivation

Also questioned is the effect of retrieval practice on student motivation, which is a key element related to efficient learning (Abel & Bäuml, 2020). According to Chen et al. (1999) intrinsic motivation (i.e., feelings of joy and interest from the activity itself apart from any extrinsic benefits or consequences) of students in an activity depends on their situational

interest. Situational interest can be defined as the appealing effect of the characteristics of an activity on individuals (Chen et al., 2001). If students have a high level of situational interest, this can increase their motivation for this activity (Deci, 1992). Retrieval practice can be seen as more engaging, appealing and active than restudying (Jing et al., 2016). This would assume that retrieval practice could increase the intrinsic motivation of students.

However, a study conducted by Abel and Bäuml (2020) showed that only retrieval practice with feedback increased the intrinsic motivation of students. In this study, the effect of retrieval practice on motivation by using a free-choice period was investigated. In this free-choice period, students had the opportunity to learn more about a foreign language in the absence of extrinsic reasons. Findings show that students are more likely to continue studying after they engaged in retrieval practice with feedback. One possible explanation for this effect is the perceived competence of students during learning, which can be defined as the confidence in the ability to learn a task (van Harsel et al., 2020). It can be argued that the perceived competence of students will decrease when they engage in retrieval practice, because this requires more mental effort (Kirk-Johnson et al., 2019). However, when students receive feedback of their performance on the retrieval practice, they gain insight in their current level of knowledge and get insight in which topics they have to focus on. Consequently, this feedback increases the experience of competence of students and thus their intrinsic motivation (Abel & Bäuml, 2020).

Present study

This study will investigate the effects of retrieval practice in instructional video-based learning on the motivation and learning of primary school students. This study has three conditions: retrieval practice between watching the instructional videos two times (RP in-between), retrieval practice after watching the instructional video two times (RP after) and only watching the instructional video without retrieval practice (restudy). The following

research question is formulated: *'Does the enrichment of instructional videos with retrieval practice increase the motivation and learning of primary school students?'*. During the experiment, the effects on mental effort are explored, because this gives insight into the efficiency of the learning process of students (van Gog & Paas, 2008). Besides, the perceived competence and metacognition of students is measured to give insight in the learning processes of students.

Retrieval practice hypothesis. First, it is hypothesized that retrieval practice will improve the long-term retention due to the effects of retrieval-based learning (Abel & Bäuml, 2020; Roediger & Butler, 2011). Thus, retrieval practice in-between or after the instructional video is expected to have a positive influence on the long-term retention and comprehension of the participants. Moreover, the effects on an immediate posttest are explored, due to the conflicting evidence of the effects of retrieval practice on short-term memory.

Timing hypothesis. Second, it is hypothesized that retrieval practice in-between watching instructional videos is more effective than retrieval practice after watching instructional videos (Lachner et al., 2019; van der Zanden, 2020). Students in the RP in-between condition get the opportunity to focus on elements which they do not master yet, which can increase their performance (Szpunar et al., 2014). Although the studies of Lachner et al. (2019) and van der Zanden (2020) incorporated only a direct posttest, it can be expected that the effects remain or are even stronger on a delayed posttest, because retrieval practice shows a strong beneficial effect on long-term memory (Roediger & Butler, 2011).

Motivation hypothesis. Third, it is hypothesized that retrieval practice is more motivating than restudying, because restudying is less engaging, appealing and active than retrieval practice (Jing et al., 2016). More specifically, it is hypothesized that the RP in-between condition, which can be seen as retrieval practice with feedback, is more motivating than the RP after condition (Abel & Bäuml, 2020).

Methods

Participants and design

The participants were 96 primary school children (52 girls, 44 boys; age: $M = 10.78$, $SD = 0.73$) in the fifth and sixth grade of three primary schools in the Netherlands. One student was excluded from the experiment, because this student withdrew during the experiment, leaving a sample of 95 students (43 boys, 52 girls; age: $M = 10.77$, $SD = 0.72$). The experiment had a between-subject design in order to compare differences between groups. Participants were randomly assigned to one of three experimental conditions: 1) RP in-between: the participants watched the video, then received retrieval practice and finished with watching the video for the second time ($N = 32$; 13 boys, 19 girls), 2) RP after: the participants watched the instructional video two times and then received retrieval practice ($N = 31$; 14 boys, 17 girls) and 3) Restudy: the participants watched the instructional video three times ($N = 32$; 16 boys, 16 girls). The experiment consisted of three phases: (a) learning phase, (b) immediate posttest, and (c) delayed posttest. At the delayed posttest, which was completed after three or four days, 22 participants were absent due to quarantine and/or illness. So, the sample for the effects on the delayed posttest consists of 73 participants (34 boys, 39 girls; age: $M = 10.82$, $SD = 0.73$), almost equally spread among the conditions: RP in-between ($N = 23$; 10 boys, 13 girls), RP after ($N = 26$; 11 boys, 15 girls) and restudy ($N = 24$; 13 boys, 11 girls).

The test power with a medium effect size for the acquired sample was .57, computed in G*Power 3.1. However, to gain a good power of $>.80$, the requested sample size with a medium effect size was $N = 159$, with a α of .05.

Materials

The experiment was presented in Qualtrics, an online-survey tool. The materials in the

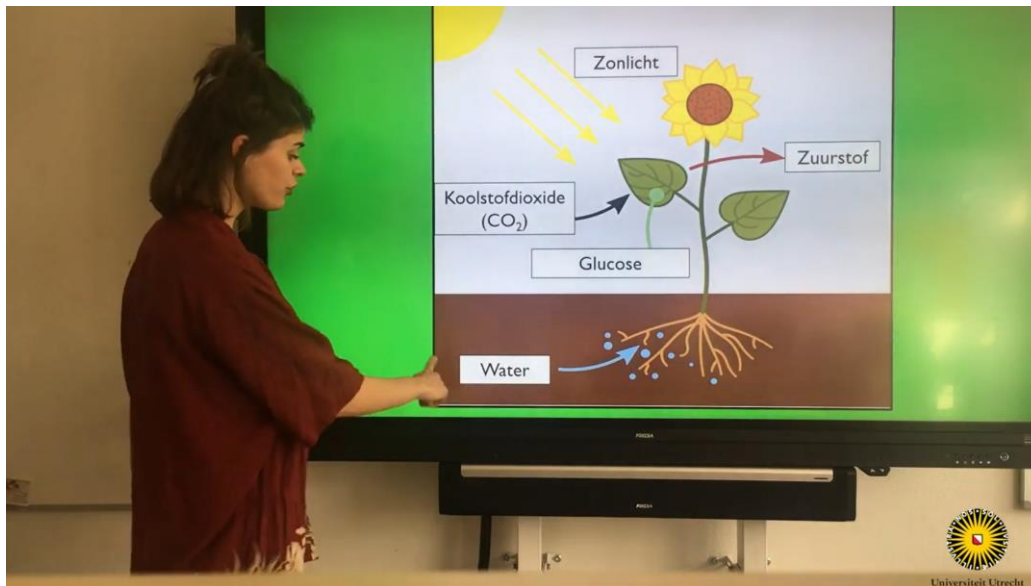
experiment were previously used by van der Zanden (2020) and Hoogerheide et al. (2019) and adapted from an existing biology method – Biology for you – used in the first grade of secondary school. Thus, participants were assumed to be novices on the subject of the instructional videos (i.e., photosynthesis) as this subject had not (yet) been part of their study program.

Self-reported prior knowledge. The prior knowledge of the participants was measured by a self-report test (Hoogerheide et al., 2019). Participants had to estimate their current knowledge of photosynthesis on a 5-point Likert scale ranging from ‘very low’ to ‘very high’ ($\alpha = .721$). Besides, students had to indicate which of the five statements applied to them, such as: ‘I know exactly why photosynthesis is important for humans’. Participants did not get an objective test which measures their prior knowledge, because an objective test could give them indications on which elements they should focus in the learning phase (Hoogerheide et al., 2019).

Instructional video. The participants watched an instructional video of six minutes and thirteen seconds¹. The video was created for the purpose of the study of van der Zanden (2020). In the video, a teacher explains the process and importance of photosynthesis for both animals and humans (figure 1). The teacher explains the characteristics of the leafs, the substances needed for photosynthesis (e.g., water) and which substances arise during photosynthesis (e.g., oxygen).

Figure 1.
Screenshot of the instructional video about photosynthesis.

¹ https://www.youtube.com/watch?v=TxemIJ1_dr4



Retrieval practice. The assignment of the retrieval practice was free recall. The participants wrote down everything they remembered (i.e., concepts and/or content) after they watched the instructional video (Lachner et al., 2019).

Posttests. The experiment consisted of an immediate and a delayed posttest (see appendix A). The immediate posttest of van der Zanden (2020) was used for both tests. The first part consisted of ten short open-ended questions to measure factual knowledge (i.e., retention) with a reliability of $\alpha = .700$ on the immediate posttest and $\alpha = .619$ on the delayed posttest. Examples of these questions are: ‘What substance that is formed by photosynthesis is food for the plant?’ or ‘In what part of the cells of a plant does photosynthesis occur?’. The second part consisted of four long open-ended question to measure conceptual knowledge (i.e., comprehension) with a reliability of $\alpha = .680$ on the immediate posttest and $\alpha = .701$ on the delayed posttest. Examples of these questions are: ‘A company want to cut down large areas with trees, why would this be bad for humans and animals?’ and ‘For diner you eat a chicken filet, can you explain what chicken fillet has to do with photosynthesis?’.

Situational interest. Situational interest was measured after the learning phase by using the situational interest scale (Rotgans & Schmidt, 2011). The scale consists of six items

and was adjusted to the topic of this study ($\alpha = .877$). Examples of these items are: 'I want to know more about the photosynthesis' and 'I enjoyed watching videos about the photosynthesis'. Participants had to score each item on a seven-point Likert scale, ranging from 'not true at all' to 'very true'. Before the learning phase, the participants indicated their interest in biology on a 7-point Likert scale to check on differences in topic interest beforehand. Their interest in biology can in fact influence their situational interest about photosynthesis, a biology topic.

Perceived competence. Perceived competence was measured after each learning task by using the perceived competence scale for learning (van Harsel et al., 2020). The scale has three items: 'I feel confident in my ability to learn how the photosynthesis works', 'I feel able to meet the challenge of performing well in understanding and explaining the photosynthesis', and 'I am capable of understanding and explaining the photosynthesis' ($\alpha = .943$). The participants have rated these items on a seven-point Likert scale of 'not at all true' to 'very true' (van Harsel et al., 2020).

Perceived mental effort. The mental effort was measured after each learning task and after the factual and conceptual knowledge questions on both posttests. Participants had to indicate the amount of mental effort they invested in the learning activity (Paas, 1992), on a nine-point Likert scale, ranging from 'very low effort', to 'very high effort'.

Judgement of learning. After the learning task, the participants indicated their judgement of learning about their expected performance on the immediate and delayed posttest in order to examine the metacognition of the participants (Lachner et al., 2019). Metacognition refers to the implicit or explicit information individuals have about their own cognition which contributes to successful learning (Tsai et al., 2018). The participants had to indicate on a seven-point Likert scale to what extent they believe they would succeed on a test (Hoogerheide et al., 2014).

Procedure

Two weeks prior to the first session of the experiment, the parents or caretakers received a consent form. Only the children with consent participated in the experiment. The experiment took place in a classroom or from home by using tablets. Before the learning phase of the experiment, the participants received a questionnaire about demographics, their self-report of their prior knowledge, their perceived competence and their interest in biology.

During the learning phase, the participants were randomly assigned to one of the three conditions. The participants in the RP after condition watched the instruction video two times and then answered the retrieval practice question. The participants in the RP in-between condition watched the instruction video for the first time, then answered the retrieval practice question and finished with watching the video for the second time. The participants in the restudy condition watched the instructional video three times. After each learning task, all participants had to indicate their mental effort and perceived competence.

After the learning phase, the participants answered the questionnaire about their situational interest and judgement of learning. Finally, the participants filled in an immediate posttest. Two or three days later, depending on the availability of the participants, the delayed posttest took place. In the end, participants were thanked for their participation and got the contact information of the researcher for questions about the experiment.

Data analysis

Averages were computed for mental effort which was measured on seven moments (mental effort invested in each learning task and the mental effort invested in answering the factual and conceptual knowledge questions on both posttests), for perceived competence which was measured on five moments (perceived competence before the learning phase, after each learning task and before the delayed posttest), for situational interest and topic interest. Besides, the total score for prior knowledge was computed.

The data was scored by the experimenter and a second encoder based on a scoring protocol developed by Van der Zanden (2020) (see appendix A). Participants could earn a maximum of 10 points for the posttest on factual knowledge, one point per correct answer. For the conceptual knowledge, participants could earn a maximum of 12 points, three points per question. Two raters scored 10% of the total amount of posttests. The intraclass correlation coefficient was .945 for the factual knowledge questions and .942 for the conceptual knowledge questions. This indicates an excellent reliability between the two raters (Koo & Li, 2015).

Results

An alpha of .05 was used for all statistical analyses. For the effect size partial η^2 (η_p^2), values $<.06$ were interpreted as a small effect, values between 0.06 and $.14$ as a medium effect and values $>.14$ as a large effect (see Cohen, 1998). Unless otherwise indicated, the analyses consisted of one-way ANOVAs with condition as between-subject factor. Additionally, a Bayesian ANOVA was conducted for each analysis with the statistics programs JASP (JASP Team, 2018). Bayesian analyses compare the evidence for the null-hypothesis (i.e., RP in-between = RP after = Restudy) relative to the alternative hypothesis (i.e., RP in-between \neq RP after \neq Restudy), which is not quantified by the conventional analyses (van den Bergh et al., 2020). A Bayes Factor (BF) indicates how much more likely the data are under one of the hypothesis. For example, a BF_{01} of 5.00 means that the data are 5 times less likely under the alternative hypothesis than the null-hypothesis (van den Bergh et al., 2020).

Preliminary analyses. First, analyses were carried out to check if the assumptions were met to conduct parametric tests (Field, 2013). If assumptions are violated, this will be reported per analysis. Before conducting the analyses, it was checked if the conditions were comparable. ANOVAs showed no differences among the conditions with regard to gender, F

(2, 92) = 0.277, $p = .759$, $\eta_p^2 = .006$, $BF_{01} = 8.40$, age, $F(2, 92) = 0.016$, $p = .984$, $\eta_p^2 < .001$, $BF_{01} = 10.34$, topic interest in biology, $F(2, 92) = 0.900$, $p = .410$, $\eta_p^2 = .019$, $BF_{01} = 5.12$, and self-reported prior knowledge, $F(2, 92) = 0.319$, $p = .728$, $\eta_p^2 = .007$, $BF_{01} = 8.12$.

Besides, there is no significant difference between the time spend to complete the experiment among the conditions, $F(2, 92) = 0.157$, $p = .855$, $\eta_p^2 = .003$, $BF_{01} = 9.27$. Additionally, differences between participants who completed the experiment and participants who did not participate on the delayed posttest were checked. If there are no significant differences between those groups, the data of the two sessions can be compared. Differences in immediate posttest performance, situational interest and perceived competence after the learning phase were checked, because it can be assumed that these variables can influence the motivation of participants to continue the experiment. These tests revealed that there are no significant differences on the immediate posttest ($t(93) = .480$, $p = .632$), situational interest ($t(93) = .903$, $p = .369$) and perceived competence ($t(93) = 1.195$, $p = .235$). Therefore, it can be assumed that the drop-out of 22 participants was random. It was decided to analyse the immediate and delayed posttest separately in order to prevent removal of useful data. Thus, the sample of 73 participants is used for the effects on the delayed posttest and the sample of 95 participants for the other effects. Table 1 shows the data of the first session and table 2 shows the data of the second session per condition.

Table 1

Means (SD) of self-reported prior knowledge (range 1-10), immediate posttest performance, mental effort (ME; range 1-9), perceived competence (PC; range 1-7), situational interest (range 1-7), topic interest (range 1-7) and metacognition (range 1-7) per condition.

| | RP In- between condition (n = 32) | RP After condition (n = 31) | Restudy condition (n = 32) |
|---|---|-----------------------------------|----------------------------------|
| Self-reported prior knowledge | 1.56 (0.84) | 1.74 (1.18) | 1.78 (1.41) |
| Immediate posttest factual knowledge questions | 4.59 (2.20) | 4.26 (2.45) | 4.72 (2.90) |
| Immediate posttest conceptual knowledge questions | 4.66 (2.34) | 4.84 (2.79) | 5.06 (2.91) |
| ME learning task 1 | 3.38 (1.90) | 3.38 (1.63) | 3.22 (2.06) |
| ME learning task 2 | 3.97 (1.73) | 3.09 (1.62) | 3.13 (2.43) |
| ME learning task 3 | 3.28 (1.76) | 3.29 (1.71) | 2.63 (2.28) |
| ME immediate posttest factual knowledge | 5.87 (1.74) | 5.35 (2.06) | 4.72 (2.16) |
| ME immediate posttest conceptual knowledge | 5.06 (1.85) | 4.42 (1.86) | 3.72 (2.51) |
| PC before learning phase | 3.65 (1.12) | 3.82 (0.98) | 3.51 (1.39) |
| PC learning task 1 | 4.96 (1.18) | 4.92 (1.13) | 4.84 (1.55) |
| PC learning task 2 | 5.10 (1.22) | 5.08 (1.34) | 5.31 (1.63) |
| PC learning task 3 | 5.07 (1.25) | 5.24 (1.38) | 5.60 (1.61) |
| Situational interest | 5.14 (1.09) | 5.05 (1.39) | 4.92 (1.66) |
| Topic interest | 5.16 (1.32) | 4.65 (1.78) | 5.13 (1.91) |
| Metacognition | 4.44 (1.24) | 4.10 (1.68) | 4.41 (1.66) |

Table 2

Means (SD) of delayed posttest performance, mental effort (ME; range 1-9), perceived competence (PC; range 1-7) and situational interest (range 1-7) per condition.

| | RP In- Between condition (n = 23) | RP After condition (n = 26) | Restudy condition (n = 24) |
|--|--|-----------------------------------|----------------------------------|
| Delayed posttest on factual knowledge | 4.48 (1.95) | 4.46 (2.42) | 4.50 (2.50) |
| Delayed posttest on conceptual knowledge | 5.33 (2.34) | 4.58 (2.79) | 4.83 (2.92) |
| ME delayed posttest factual knowledge | 5.04 (2.53) | 4.00 (2.53) | 3.87 (2.86) |
| ME delayed posttest conceptual knowledge | 4.74 (2.43) | 3.92 (2.70) | 4.04 (3.56) |
| PC before delayed posttest | 4.67 (1.33) | 4.56 (1.39) | 5.36 (1.75) |

Posttest performance. Regarding posttest performance of the RP after condition, the Shapiro-Wilks test of normality was significant, suggesting a violation of the assumption of normality. However, given the values of kurtosis and skewness, between -1.96 and 1.96 and $N \approx 30$, it can be assumed that the data is sufficient normally distributed to conduct parametric

tests (Field, 2013).

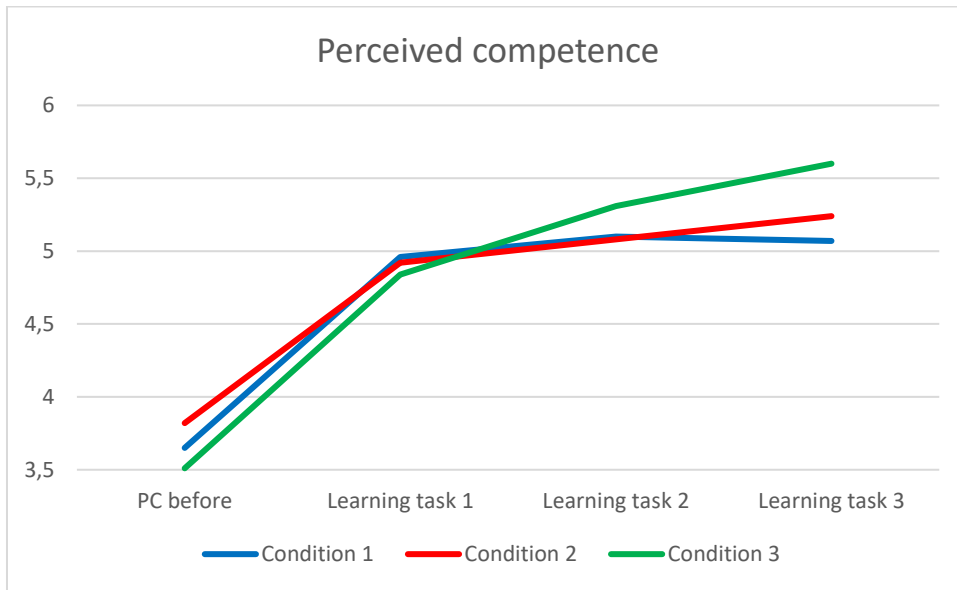
The participants could obtain a maximum of 22 points on the posttest. The participants scored an average of 9.38 points on the immediate posttest and 9.30 points on the delayed posttest, indicating an overall average score. Results showed no significant difference among the conditions on comprehension (i.e., performance on the conceptual knowledge questions) of the immediate posttest, $F(2, 92) = 0.183, p = .833, \eta_p^2 = .004, BF_{01} = 9.05$, or retention (i.e., performance on the factual knowledge questions), $F(2, 92) = 0.278, p = .758, \eta_p^2 = .006, BF_{01} = 8.40$. It was hypothesized that retrieval practice would improve long-term retention, and that retrieval practice in-between watching the instructional video would be more effective than retrieval practice after watching the instructional video. However, there was no significant difference between the conditions on comprehension (i.e., performance on the conceptual knowledge questions) of the delayed posttest, $F(2, 70) = 0.345, p = .710, \eta_p^2 = .010, BF_{01} = 6.58$, or retention (i.e., performance on the factual knowledge questions), $F(2, 70) = 0.002, p = .998, \eta_p^2 < .001, BF_{01} = 8.55$. Both the retrieval practice hypothesis and the timing hypothesis can be rejected.

Situational interest. Regarding situational interest of the RP after and restudy condition, the Shapiro-Wilks test of normality was significant, suggesting a violation of the assumption of normality. However, it was decided to conduct a parametric test because of the values of kurtosis and skewness between -1.96 and 1.96 and $N \approx 30$. It was hypothesized that retrieval practice would be more motivating than restudying, and that retrieval practice in-between watching instructional videos would be more motivating than retrieval practice after watching instructional videos. The analysis showed no significant effect on situational interest after the learning phase among the three condition, $F(2, 92) = 0.180, p = .836, \eta_p^2 = .004, BF_{01} = 9.08$. This indicates there is no significant difference in motivation between the participants. Thus, the third hypothesis of this study, the motivation hypothesis, can also be

rejected.

Perceived competence. Perceived competence was measured before the learning phase, after each learning task and before the delayed posttest. Figure 2 shows the perceived competence over time until the immediate posttest. A mixed ANOVA was carried out to check differences in perceived competence over time among the conditions until the immediate posttest (i.e., before the learning phase, learning task 1, learning task 2 and learning task 3). Mauchly's test showed that the assumption of sphericity was violated, $\chi(5) = 111.98, p < .001$, therefore the degrees of freedom are corrected with Greenhouse-Geisser ($\epsilon = .59$). Results showed no significant difference between perceived competence and condition, $F(3.54, 6.38) = 1.691, p = .162, \eta_p^2 = .036, BF_{01} = 29.83$. There was no main effect for condition, $F(2, 92) = 0.890, p = .915, \eta_p^2 = .019, BF_{01} = 6.38$. However, there was a main effect for perceived competence over time, $F(1.77, 160.79) = 85.17, p < .001, \eta_p^2 = .012, BF_{01} = 1.00$. So, the perceived competence of the participants significantly increased during the first part of the experiment. More specifically, the Bonferroni post-hoc analysis, which corrects the α of 0.05 by the number of tests conducted in order to decrease the chance of rejecting a true null hypothesis (Field, 2013), showed that the perceived competence after the first learning task ($M = 4.91, SD = 1.29$) was significantly higher than the perceived competence before the learning phase ($M = 3.66, SD = 1.17$), $p < .001, d = 1.016$. Besides this, the perceived competence after the second learning task ($M = 5.17, SD = 1.40$), $p = .001, d = 0.191$. was significantly higher than after the first learning task. The perceived competence after learning task 3 ($M = 5.03, SD = 1.42$) was not significantly higher than the perceived competence after learning task 2, $p = .227, d = 0.093$. Additionally, there was no significant effect on the perceived competence before the delayed posttest among the conditions, $F(2, 70) = 2.031, p = .139, \eta_p^2 = .055, BF_{01} = 1.83$.

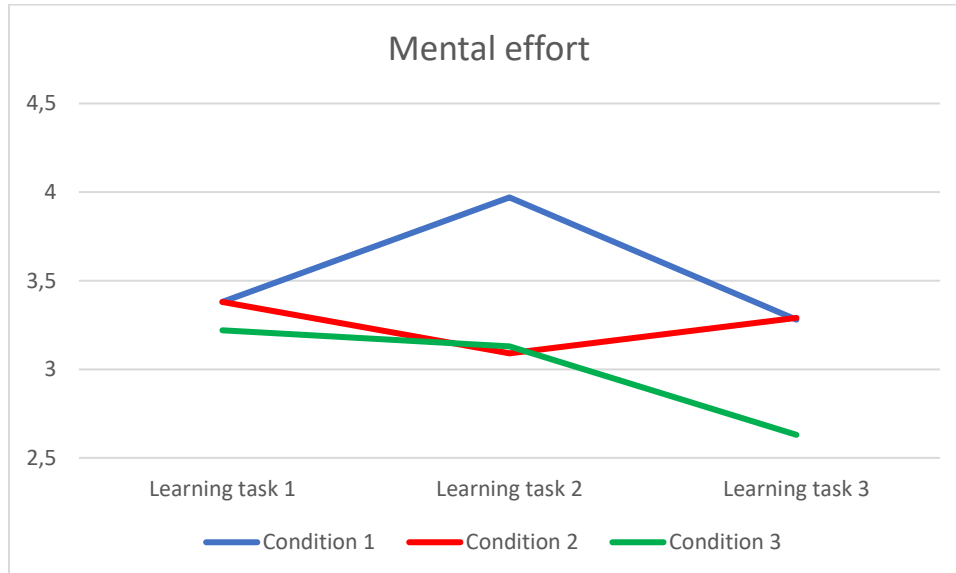
Figure 2.
Mean scores on perceived competence (range 1-7).



Mental effort. Mental effort was measured after each learning task (i.e., three moments during the learning phase) and after the immediate and delayed posttest (i.e., four times). Figure 3 shows the mental effort for each condition over time during the learning phase. A mixed ANOVA was carried out to check differences in mental effort during the learning phase (e.g., learning task 1, 2 and 3). Results showed no significant difference of mental effort and condition, $F(4, 184) = 1.728, p = .146, \eta_p^2 = .038, BF_{01} = 18.01$. There was no main effect for condition, $F(2, 92) = .915, p = .404, \eta_p^2 = 0.20, BF_{01} = 3.68$ and no main effect for mental effort $F(2, 184) = 1.851, p = .160, \eta_p^2 = .020, BF_{01} = 4.91$. Besides this, the mental effort invested in the retrieval practice between condition 1 and 2 was compared. Results showed no significant difference in mental effort of the retrieval practice condition between the conditions $F(1, 61) = 2.438, p = .124, \eta_p^2 = .038, BF_{01} = 1.40$. In order to check difference in mental effort on the posttests, multiple ANOVAs were carried out. Results of the immediate posttest showed no significant difference among the conditions on their mental effort used while answering the factual knowledge questions, $F(2, 92) = 2.704, p = .072, \eta_p^2 = .056, BF_{01} = 1.22$. However, the conditions significantly differ on their mental effort on the

conceptual knowledge questions, $F(2, 92) = 3.292, p = .042, \eta_p^2 = .067, BF_{01} = 1.30$. Results obtained by conducting Bonferroni post-hoc analysis showed that the mean score for the RP in-between condition ($M = 5.06, SD = 1.85$) was significantly higher than the restudy condition ($M = 4.42, SD = 1.86$), $p = .036, d = .609$. The RP-after condition ($M = 3.72, SD = 2.51$) did not significantly differ from the from the RP in-between condition, $p = .621, d = .345$, nor the restudy condition, $p = .564, d = .317$. This indicates that participants in the restudy condition reported a lower mental effort while answering the conceptual knowledge questions on the immediate posttest than participants in the RP in-between condition. Results of the delayed posttest showed no significant difference among the conditions on the factual knowledge questions, $F(2, 70) = 1.385, p = .257, \eta_p^2 = .038, BF_{01} = 2.99$, nor on the conceptual knowledge questions, $F(2, 70) = .537, p = .587, \eta_p^2 = .015, BF_{01} = 5.69$.

Figure 3.
Mean scores on mental effort (range 1-9).



Judgement of learning. The participants indicated their judgement of learning about their expected performance on the immediate and delayed posttest. Results showed no significant differences among the conditions on their judgements of learning for the immediate posttest, $F(2, 92) = .468, p = .627, \eta_p^2 = .010, BF_{01} = 7.22$, or the delayed posttest,

$F(2, 70) = 2.911, p = .061, \eta_p^2 = .077, BF_{01} = 1.06$. In short, there is no difference in the judgement about the expected performance on both posttest among the conditions.

Discussion

The aim of the present study was to investigate whether retrieval practice in-between or after video-based learning enhances the comprehension and retention on immediate and delayed posttest performance and increases the motivation of primary school students. Prior research on retrieval practice shows promising evidence for the long-term memory of students (Roediger & Butler, 2011). However, evidence regarding the effect of retrieval practice on video-based learning and motivation is quite small. It is important to extend our knowledge about video-based learning, because instructional videos are nowadays one of the most popular ways to deliver instruction. This research attempted to clarify this research gap by replicating and extending the study of van der Zanden (2020).

The results of this study indicate that retrieval practice in-between or retrieval practice after studying instructional videos does not increase the comprehension or retention on an immediate posttest. In other words, this study failed to replicate the effect found by van der Zanden (2020), who argued that retrieval practice increases retention on an immediate posttest. One possible explanation is the different set-up of the experiment, because the time-on-task was kept equal in the present study. The participants in the restudy condition spend the same amount of time on the task, which can influence their performance on the posttest. Additionally, the participants in the retrieval practice in-between condition reported a higher amount of mental effort while answering the conceptual knowledge questions compared to the restudy condition. According to Van Gog and Paas (2008), activities which require more mental effort, due to a higher amount of cognitive load, result in a higher performance and lower mental effort on the test. So, the higher amount of mental effort on the test in this study could indicate that the participants did not put in enough effort to answer the retrieval practice

question. In this case, the participants in the in-between retrieval practice condition could have more difficulties on the immediate posttest because they failed to put in enough effort during the learning phase. However, the result of the present study regarding the performance on the immediate posttest is in line with previous research (Roediger & Butler, 2011; Roediger & Karpicke, 2006). Bayesian analyses showed strong evidence for the null-hypothesis on immediate posttest performance (i.e., Bayes Factors of 9.05 and 8.40).

Surprisingly, this study could not find support for the retrieval practice hypothesis which expected that retrieval practices improves long-term retention. The results indicate that retrieval practice in video-based learning does not increase the comprehension or retention on a delayed posttest which contradicts previous research (see Roediger and Butler, 2011). According to the ICAP framework, it could be expected that retrieval practice, which is an active activity, leads to better learning outcomes (Chi & Wylie, 2014). However, Bayesian analyses showed strong evidence for the null-hypothesis on delayed test performance (i.e., Bayes Factors of 6.58 and 8.55; Jarosz & Wiley, 2014). Previous research shows that the effect of retrieval practice on long-term memory is quite robust, especially for factual retention (Roediger & Butler, 2011). It seems the participants failed to answer the free recall question in depth (e.g., the average answer was only 34 words from a six minute video), which could indicate that the criterion level could be too low. In other words, it could be possible that the participants did not have enough knowledge to recall during the free recall question. According to Kornell et al. (2011) the benefits of retrieval practice depend on the success of the retrieval practice; restudying is more effective than unsuccessful testing, but not as effective as successful testing.

Besides, this study only wanted to investigate the performance on a delayed posttest. Due to the global COVID-19 pandemic, it was expected that the dropout of students could be high. Thus, it was decided to incorporate an immediate posttest in order to prevent the

removal of important data because of the absence during the delayed posttest. Additionally, there is conflicting evidence about the effects of retrieval practice on immediate performance, so this study wanted to explore this effect. Due to this decision, all the participants engaged in at least one retrieval practice activity (e.g., free recall and/or the immediate posttest) before the delayed posttest. The immediate posttest can in fact be seen as retrieval practice without feedback, because students actively retrieve and reconstruct information from the videos during the immediate posttest. This means that the restudy condition also engaged in retrieval practice before the delayed posttest which can influence the performance.

Additionally, there was no significant support found for the timing hypothesis, indicating there is no significant difference on the posttest performance between retrieval practice in-between and after watching the instructional videos which contradicts the findings of Lachner et al. (2019) and van der Zanden (2020). It could be expected that retrieval practice in-between watching instructional videos would be more effective, because it can give students a metacognitive advantage. Retrieval practice during the study phase gives students awareness of their knowledge gaps early on. It should increase their attention and they can invest more effort in restudying the materials because of their awareness of their current knowledge (Lachner et al., 2019). However, the students in the retrieval practice in-between condition did not put more effort in answering the retrieval practice question or during the second time they watched the video.

This study found a positive significant effect of perceived competence over time, which indicates that the participants gained more confidence about their ability to learn about photosynthesis during the learning phase. Surprisingly, there was no significant difference among the conditions on perceived competence. According to Kirk-Johnson et al. (2019), perceived competence will decrease after students engage in retrieval practice, because this requires more mental effort. However, this was not the case in the present study, because the

students indicated that they did not put more mental effort in answering the retrieval practice question than in watching the videos. Thus, this is a possible explanation why the students also did not report a lower perceived competence after retrieval practice.

Although retrieval practice can be seen as more appealing and engaging, this study did not find support for the motivation hypothesis, which hypothesized that retrieval practice is more motivating than restudying. Bayes Factor of 9.08 indicates strong evidence for the null-hypothesis on situational interest (Jarosz & Wiley, 2014). One possible explanation for this conflicting evidence could be the excitement of the participants to participate with a research for the first time. This excitement and curiosity to participate in a research could influence their situational interest for the learning activities, which was relatively high for all the conditions, 5.04 on a 7-point Likert scale, even if they only had to restudy the materials.

Additionally, there is no difference in motivation between retrieval practice with feedback and retrieval practice without feedback. Abel and Bäuml (2020) argued that only retrieval practice with feedback increases the motivation of students. It was argued that retrieval practice in-between watching instructional videos can be seen as retrieval practice with feedback, because participants get insight in their current knowledge and knowledge gaps. However, taken into account the development of executing effective learning strategies of primary school students (see Pressley & Harris, 2009), it can be expected that they are not capable of gaining insight in their current knowledge by restudying the video after retrieval practice. In this case, retrieval practice in-between can not be considered as retrieval practice with feedback anymore, which could be a possible explanation of this finding.

Aside from the limitation that the restudy condition also participated in retrieval practice because of the incorporation of the immediate posttest, this study has some other limitations. First of all, this study has a medium test power of .57. In order to gain a good test power, the number of participants should be $N = 159$. The implication of a medium test power

is that there is a small chance to find a true effect (Bryman, 2015). It is more likely that you find support for the null hypothesis, while you actually should reject this hypothesis (Bryman, 2015). With a bigger sample size, it could be possible to detect some significant effects. However, the Bayesian statistics of this study show a really strong effect supporting the hypothesis. Thus, it can be questioned if a bigger sample size could increase the chance to reject the null hypothesis. Second, this study did not score the free recall activity which could give more insight in the performance of the retrieval practice. Only the amount of words were calculated to get some insight. Third, the participants of this study were randomly assigned over the three conditions, so it could be expected that the average level of students is equal among the conditions. However, there are some huge differences in levels of students at primary school, which could bias the results.

In conclusion, this study did not find evidence for the benefits of retrieval practice on retention and comprehension of an instructional video about a biology topic with primary school students on their long-term memory. However, this study showed that the effects of retrieval practice are not noticeable on an immediate posttest, which is in line with previous research. Given the lack of power and the fact this study did not find an effect of retrieval practice on long-term memory which contradicts previous research, it's important to be careful when drawing firm conclusions. Besides, there are still many unanswered questions about the influence of retrieval practice on student motivation. Thus, future research is needed to investigate whether these effects can be replicated and generalized to other materials and populations.

Reference list

- Abel, M., & Bäuml, K.-H. T. (2020). Would you like to learn more? Retrieval practice plus feedback can increase motivation to keep on studying. *Cognition*, *201*, 104316
<https://doi.org/10.1016/j.cognition.2020.104316>
- Behehsti, M., Taspolat, A., Kaya, M. S., & Sapanca, H. F. (2018). Characteristics of instructional videos. *World Journal on Educational Technology: Current Issues*, *10*(1), 61-69. <https://doi.org/10.18844/wjet.v10i1.3186>
- Bryman, A. (2015). *Social Research Methods* (5de ed.). Oxford University Press.
- Chen, A., Darst, P. W., & Pangrazi, R. P. (1999). What constitutes situational interest? validating a construct in physical education. *Measurement in Physical Education and Exercise Science*, *3*(3), 157–180. https://doi.org/10.1207/s15327841mpee0303_3
- Chen, A., Darst, P. W., & Pangrazi, R. P. (2001). An examination of situational interest and its sources. *British Journal of Educational Psychology*, *71*(3), 383–400.
<https://doi.org/10.1348/000709901158578>
- Chi, M. T. H. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, *1*(1), 73–105.
<https://doi.org/10.1111/j.1756-8765.2008.01005.x>
- Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, *49*(4), 219–243.
<https://doi.org/10.1080/00461520.2014.965823>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

- Cummins, S., Beresford, A. R., & Rice, A. (2016). Investigating engagement with in-video quiz questions in a programming course. *IEEE Transactions on Learning Technologies*, 9(1), 57–66. <https://doi.org/10.1109/tlt.2015.2444374>
- De Koning, B. B., Hoogerheide, V., & Boucheix, J. M. (2018). Developments and trends in learning with instructional video. *Computers in Human Behavior*, 89, 395-398. <https://doi.org/10.1016/j.chb.2018.08.055>
- Deci, E. L. (1992). Article commentary: On the nature and functions of motivation theories. *Psychological Science*, 3(3), 167–171. <https://doi.org/10.1111/j.1467-9280.1992.tb00020.x>
- Earp, B. D., & Trafimow, D. (2015). Replication, falsification, and the crisis of confidence in social psychology. *Frontiers in Psychology*, 6 (1), 1-11. <https://doi.org/10.3389/fpsyg.2015.00621>
- Field, A. (2013). *Discovering statistics using SPSS* (4th ed.). Sage publications.
- Goossens, N. A. M. C., Camp, G., Verkoeijen, P. P. J. L., & Tabbers, H. K. (2013). The effect of retrieval practice in primary school vocabulary learning. *Applied Cognitive Psychology*, 28(1), 135–142. <https://doi.org/10.1002/acp.2956>
- Hoogerheide, V., Loyens, S. M. M., & van Gog, T. (2014). Comparing the effects of worked examples and modeling examples on learning. *Computers in Human Behavior*, 41, 80-91. <https://doi.org/10.1016/j.chb.2014.09.013>
- Hoogerheide, V., Visee, J., Lachner, A., & van Gog, T. (2019). Generating an instructional video as homework activity is both effective and enjoyable. *Learning and Instruction*, 64, 101226. <https://doi.org/10.1016/j.learninstruc.2019.101226>

- Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting bayes factors. *The Journal of Problem Solving*, 7(1).
<https://doi.org/10.7771/1932-6246.1167>
- Jing, H. G., Szpunar, K. K., & Schacter, D. L. (2016). Interpolated testing influences focused attention and improves integration of information during a video-recorded lecture. *Journal of Experimental Psychology: Applied*, 22(3), 305–318.
<https://doi.org/10.1037/xap0000087>
- Karpicke, J. D. (2012). Retrieval-Based Learning. *Current Directions in Psychological Science*, 21(3), 157–163. <https://doi.org/10.1177/0963721412443552>
- Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, 331, 772–775.
<https://doi.org/10.1126/science.1199327>
- Kirk-Johnson, A., Galla, B. M., & Fraundorf, S. H. (2019). Perceiving effort as poor learning: The misinterpreted-effort hypothesis of how experienced effort and perceived learning relate to study strategy choice. *Cognitive Psychology*, 115, 101237.
<https://doi.org/10.1016/j.cogpsych.2019.101237>
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155–163.
<https://doi.org/10.1016/j.jcm.2016.02.012>
- Kornell, N., Bjork, R. A., & Garcia, M. A. (2011). Why tests appear to prevent forgetting: A distribution-based bifurcation model. *Journal of Memory and Language*, 65(2), 85–97.
<https://doi.org/10.1016/j.jml.2011.04.002>

- Lachner, A., Backfisch, I., Hoogerheide, V., van Gog, T., & Renkl, A. (2019). Timing matters! Explaining between study phases enhances students' learning. *Journal of Educational Psychology*, *112*(4), 841–853. <https://doi.org/10.1037/edu0000396>
- Leisner, D., Zahn, C., Ruf, A., & Cattaneo, A. (2020). Different ways of interacting with videos during learning in secondary physics lessons. *Communications in Computer and Information Science*, *1225*, 284–291. https://doi.org/10.1007/978-3-030-507299_40
- Moreira, B. F. T., Pinto, T. S. S., Justi, F. R. R., & Jaeger, A. (2019). Retrieval practice improves learning in children with diverse visual word recognition skills. *Memory*, *27*(10), 1423–1437. <https://doi.org/10.1080/09658211.2019.1668017>
- Paas, F. G. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of educational psychology*, *84*(4), 429.
- Pressley, M., & Harris, K. R. (2009). Cognitive strategies instruction: from basic research to classroom instruction. *Journal of Education*, *189* (2), 77–94. <https://doi.org/10.1177/0022057409189001-206>
- Roediger III, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long term retention. *Trends in Cognitive Sciences*, *15*(1), 20–27 <https://doi.org/10.1016/j.tics.2010.09.003>
- Roediger III, H. L., & Karpicke, J. D. (2006). Test-enhanced learning. *Psychological Science*, *17*(3), 249 – 255. <https://doi.org/10.1111/j.1467-9280.2006.01693.x>

Rotgans, J. I., & Schmidt, H. G. (2011). Situational interest and academic achievement in the active-learning classroom. *Learning and Instruction, 21*(1), 58-67.

<https://doi.org/10.1016/j.learninstruc.2009.11.001>

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin, 140*(6), 1432-1463.

<https://doi.org/10.1037/a0037559>

Szpunar, K. K., Jing, H. G., & Schacter, D. L. (2014). Overcoming overconfidence in learning from video-recorded lectures: Implications of interpolated testing for online education. *Journal of Applied Research in Memory and Cognition, 3*(3), 161–164.

<https://doi.org/10.1016/j.jarmac.2014.02.001>

Szpunar, K. K., Khan, N. Y., & Schacter, D. L. (2013). Interpolated memory tests reduce mind wandering and improve learning of online lectures. *Proceedings of the National Academy of Sciences, 110*(16), 6313–6317. <https://doi.org/10.1073/pnas.1221764110>

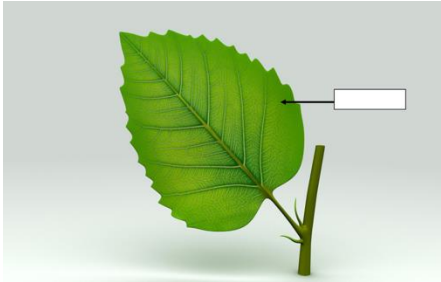
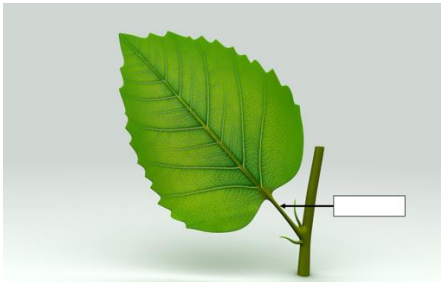
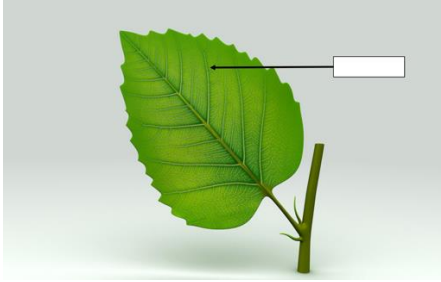
Tsai, Y. H., Lin, C. H., Hong, J. C., & Tai, K. H. (2018). The effects of metacognition on online learning interest and continuance to learn with MOOCs. *Computers & Education, 121*(1), 18–29. <https://doi.org/10.1016/j.compedu.2018.02.011>

van den Bergh, D., van Doorn, J., Marsman, M., Draws, T., van Kesteren, E. J., Derks, K., Dablander, F., Gronau, Q. F., Kucharský, I., Gupta, A. R. K. N., Sarafoglou, A., Voelkel, J. G., Stefan, A., Ly, A., Hinne, M., Matzke, D., & Wagenmakers, E. J. (2020). A Tutorial on Conducting and Interpreting a Bayesian ANOVA in JASP. *L'Année psychologique, 120*(1), 73. <https://doi.org/10.3917/anpsy1.201.0073>

- Van den Broek, G. S. E., Segers, E., Takashima, A., & Verhoeven, L. (2013). Do testing effects change over time? Insights from immediate and delayed retrieval speed. *Memory*, 22(7), 803–812. <https://doi.org/10.1080/09658211.2013.831455>
- Van Gog, T., & Paas, F. (2008). Instructional efficiency: revisiting the original construct in educational research. *Educational Psychologist*, 43(1), 16–26. <https://doi.org/10.1080/00461520701756248>
- Van Harsel, M., Hoogerheide, V., Verkoeijen, P., & van Gog, T. (2020). Examples, practice problems, or both? Effects on motivation and learning in shorter and longer sequences. *Applied Cognitive Psychology*, 34(4), 793–812. <https://doi.org/10.1002/acp.3649>
- Van der Zanden, T. (2019). *Integrating retrieval practice between learning activities in an instructional video-based learning environment enhances primary school students' performance on factual retention.*
- Van der Zee, T., Davis, D., Saab, N., Giesbers, B., Ginn, J., van der Sluis, F., Paas, F., & Admiraal, W. (2018). Evaluating retrieval practice in a MOOC. Proceedings of the 8th International Conference on Learning Analytics and Knowledge, 216–225. <https://doi.org/10.1145/3170358.3170382>
- Yong, P. Z., & Lim, S. W. H. (2016). Observing the testing effect using Coursera video-recorded lectures: A preliminary study. *Frontiers in Psychology*, 6. <https://doi.org/10.3389/fpsyg.2015.02064>

Appendix A

Factual knowledge items

| No. | Question | Answer | Scoring |
|-----|---|-----------|---------|
| 1 | Below you see the image of a leaf, which component of the leaf is pointed out by the arrow? | Bladmoes | 1 point |
| |  | | |
| 2 | Below you see the image of a leaf, which component of the leaf is pointed out by the arrow? | Bladsteel | 1 point |
| |  | | |
| 3 | Below you see the image of a leaf, which component of the leaf is pointed out by the arrow? | Zijnerf | 1 point |
| |  | | |

| | | | |
|----|---|---|---------|
| 4 | People feed themselves by eating other organisms, how do plants feed themselves? | Photosynthesis | 1 point |
| 5 | What component for the process of photosynthesis does a plant absorb from the ground? | Water | 1 point |
| 6 | What component for the process of photosynthesis can you also find in soda (e.g. cola) | Carbon dioxide or Co ₂ | 1 point |
| 7 | For what component of photosynthesis, it is important that a plant does not stand in the shadow for too long? | (Sun)light | 1 point |
| 8 | What substance that is formed by photosynthesis is food for the plant? | Glucoses | 1 point |
| 9 | What substance that is formed by photosynthesis is important for humans and animals? | Oxygen or O ₂ | 1 point |
| 10 | In what part of the cells of a plant does photosynthesis occur? | Chloroplast or All green parts | 1 point |

Factual knowledge items

| No. | Question | Answer + scoring |
|-----|---|---|
| 11 | At breakfast you eat a sandwich, can you explain what a sandwich has to do with photosynthesis? | - Bread is made of wheat (1p) - What grows on plants (1p) - Plants need photosynthesis (1p) |

- 12 For diner you eat a chicken fillet, can you explain what chicken fillet has to do with photosynthesis?
- Fillet comes from a chicken (1p)
 - Chicken eats plants (1p)
 - Plants need photosynthesis (1p)
- 13 A company wants to cut down large areas with trees, why would this be bad for humans and animals?
- Trees produces oxygen (1p)
 - Humans and animals need oxygen to live (1p)
 - Without this humans and animals can't live (1p)
- 14 Before the schools closed there was a beautiful plant in the classroom, when the schools re-opened the leafs are brown and are hanging down, what happened?
- Plant had no sunlight or water (1p)
 - Sunlight or water is needed for photosynthesis (1p)
 - Without photosynthesis a plant has no food (1p)
-