Is two better than one? Effects of sequencing different generative learning strategies on learning from instructional videos.

Suzanne van Laarschot

6692818

Department of Social Sciences, Utrecht University

Master's Thesis

First assessor: dr. Vincent Hoogerheide

Second assessor: dr. Liesbeth Kester

Word count: 7994

Abstract

Learners can adopt generative learning strategies to transform learning from instructional videos from a passive to an active learning experience, which has proven to increase learning. However, it remains an open question if enriching videos with multiple generative learning strategies further enhances learning and if so, how these strategies should best be sequenced. The purpose of the present study was to examine whether engaging in multiple generative learning strategies (self-explanation and retrieval practice) rather than a single strategy would increase learning from instructional videos. Additionally, this study examined whether the sequence of self-explanation and retrieval practice (SE-RP vs. RP-SE) affected learning. A between-subjects design was used where 155 Dutch secondary vocational education students were randomly assigned to a RP-RP, SE-SE, RP-SE, or SE-RP condition. Participants watched an instructional video and subsequently completed a retrieval practice or self-explanation task. After completing the task, participants watched the video again and completed another retrieval practice or self-explanation task. After approximately one week, 95 participants completed a delayed posttest that measured their retention, comprehension, and transfer. Findings showed no significant effects of the use multiple generative learning strategies or sequencing of retrieval practice and self-explaining on students' learning outcomes.

Keywords: generative learning, video learning, sequencing, retrieval practice, self-explaining

Is two better than one? Effects of sequencing different generative learning strategies on learning from instructional videos.

Based on constructivism, learning can be considered a generative activity where learners actively construct meaning from the learning material and integrate this new information with their prior knowledge (Fiorella & Mayer, 2016; Wittrock, 2010). To foster generative learning, learners can adopt *generative learning strategies* which consist of various activities that stimulate learners to make sense of the learning material and go beyond the presented information (Brod, 2020). Previous studies on generative learning emphasized the need for future research to explore the effects of sequencing different generative learning strategies (Fiorella & Mayer, 2016; Lachner et al., 2021). Further investigation is important to enhance the effectiveness of generative learning strategies by understanding how different strategies affect learning in particular phases of knowledge acquisition and to explore interactions among strategies across different contexts (Fiorella & Mayer, 2016; Lachner et al., 2021).

Two of the most used generative learning strategies are *self-explanation* and *retrieval practice*, both have been well-studied and demonstrated to positively affect learning (e.g., Bisra et al., 2018; Karpicke et al., 2014). However, it remains an open question if using both of these generative learning strategies further increases learning, since self-explanation and retrieval practice engage learners in different forms of cognitive processing. Yet, the amount of research on combining these generative learning strategies remains scarce (Larsen et al., 2013; Roelle & Nückles, 2019). Therefore, this study adheres to the previously mentioned call for additional research on sequencing different generative learning strategies, which may contribute to the extension of generative learning theory with regards to the effects and use of multiple learning strategies. Furthermore, this study may impact practice by providing educators with practical implications on how to use multiple generative learning strategies to enhance learning.

This study examined the effects of sequencing generative learning strategies in a video learning context, since the use and popularity of instructional videos in education grew rapidly over the past years (De Koning et al., 2018; Poquet et al., 2018). Additionally, incorporating a generative learning strategy transforms learning from instructional videos from a passive to an active learning experience, which has proven to foster the effectiveness of video learning (e.g., Fiorella et al., 2020; Kleiman et al., 2019; Mayer et al., 2020). However, it remains an open question if enriching videos with multiple generative learning strategies further enhances learning and if so, how these strategies should best be sequenced when learning from videos.

The next paragraph first describes the effects of retrieval practice and self-explanation on learning and the theory behind these strategies. Additionally, the use of multiple generative strategies and their sequence will be discussed.

Effects of Retrieval Practice or Self-Explaining on Learning Outcomes

Retrieval practice refers to the process of deliberately recalling information from memory (Roediger & Butler, 2011). Previous research on retrieval practice has shown that actively retrieving information from memory enhances long-term memory retention and knowledge transfer (e.g., Bae et al., 2019; Butler, 2010; Roediger & Butler, 2011). Furthermore, meta-analysis revealed a medium effect size ($g = 0.50$, $p < .001$) of retrieval practice on learning (Rowland, 2014). Moreover, retrieval practice not only promotes retention of the practiced material itself, but also enhances learning of subsequent materials (Kliegl & Bäuml, 2021). Two types of retrieval practice tasks can be distinguished: targeted retrieval (specific short-answer tasks) and holistic retrieval (unspecific free-recall tasks). Being more specific, targeted retrieval tasks lead to greater retention of targeted information from the learning material, whereas free-recall tasks lead to better retention of non-targeted information relevant to the learning material (Endres et al., 2020). Lastly, retrieval practice can be performed in oral or written form, however there is no evidence that the effectiveness differs between these retrieval practice modes (Putnam & Roediger, 2013).

Additionally, while the amount of studies on retrieval practice in learning from instructional videos remains limited, previous research in this domain has shown promising results (Kleiman et al., 2019; Van der Zee et al., 2018). For example, Kleiman et al. (2019) conducted an experiment where 40 medical students were assigned to either a retrieval practice condition or a standard practice condition while watching several video clips. Results indicated that students who engaged in retrieval practice showed significant better long-term retention and recall than the standard practice group (Kleiman et al., 2019).

Theoretical explanations of retrieval practice mainly describe the effects on human memory. Retrieving information from long-term memory may foster learning through a consolidation function, by strengthening memory traces which enhances the accessibility of the targeted knowledge (Carpenter, 2009; Roediger & Butler, 2011). According to *elaborative retrieval theory*, retrieval practice activates related knowledge in semantic memory, which supports the development of new retrieval cues that serve as additional memory traces for the targeted knowledge (Carpenter, 2009; Endres & Renkl, 2015). Furthermore, *retrieval effort theories* suggest that the difficulty or effort induced by retrieval practice positively affects reprocessing of memory traces, with more effort leading to stronger memory traces (Pyc & Rawson, 2009; Rowland, 2014). The concept of retrieval effort might also explain why retrieval practice seems more effective in delayed testing than in immediate testing after initial study, as learners have to exert more effort to retrieve information with delayed testing (Roediger & Butler, 2011). However, an important boundary condition for retrieval practice is the amount of successful retrieval, since the benefits of retrieval practice remain limited when learners cannot successfully retrieve the targeted information (Kang et al., 2007). Additionally, repeatedly using the same retrieval practice episode induces better memory retention than a single retrieval practice session (Latimier et al., 2021).

A strategy that includes some retrieval practice, yet goes further than merely retrieving information is self-explanation. Self-explaining is defined as generating explanations to oneself based on both the given information as well as inferences that go beyond this information (Rittle-Johnson & Loehr, 2017). A recent meta-analysis on self-explanation has shown a medium effect size of self-explaining on learning outcomes (g = 0.55, p < .001) across a range of subject domains (Bisra et al., 2018). Research on self-explaining while learning from videos has shown similar positive effects of self-explanation on learning (Fiorella et al., 2020; Lawson & Mayer, 2021; Pi et al., 2021). For example, Fiorella et al. (2020) conducted an experiment where 196 college students watched a video lesson and engaged in either self-explaining, drawing, or rewatching after each part of the lesson. Significant medium to large effect sizes were found for self-explaining on retention and transfer compared to the other conditions (Fiorella et al., 2020).

The cognitive processes behind self-explanation not only consist of retrieval practice, but also of *knowledge organization* and *knowledge integration* (Fiorella & Mayer, 2015). The amount of retrieval practice while self-explaining depends on the availability of the learning materials while explaining. When learners do not have access to the learning materials, they must first retrieve the material from memory before they can explain it to oneself (Hiller et al., 2020). Furthermore, by self-explaining learners integrate information within the learning material and organize this new information by connecting it with their prior knowledge and existing mental models (Chi, 2000; Roy & Chi, 2005). Consequently, self-explanation aids learners in recognizing discrepancies between new information and their prior knowledge in order to revise inaccurate mental models (Wylie & Chi, 2014).

Several factors influence the effectiveness of self-explaining. Some studies have shown that the beneficial effects of self-explaining may be enhanced when students receive training on how to use the self-explanation strategy effectively (Dunlosky et al., 2013). Moreover, the modality of the explanations (written vs. oral) affects students' knowledge acquisition. Specifically, written explanations support the process of knowledge organization and, in turn, students' conceptual knowledge and oral explanations promote students' elaborative processes and transfer of knowledge (Lachner et al., 2018). Furthermore, the timing of self-explaining during the study phase affects its effectiveness. A study by Lachner et al. (2020) examined whether the timing of explaining affected students' learning and found that explaining early in the study phase improved learning compared to explaining at the end of the study phase. A possible explanation for this finding is that explaining early in the study phase enables learners to recognize potential knowledge gaps, which they can repair during the remaining study phase (Lachner et al., 2020).

In sum, research has shown that retrieval practice mainly benefits long-term memory retention through a consolidation function. While self-explanation goes beyond mere retrieval and additionally aids comprehension and transfer by integrating and organizing the learning material with existing prior knowledge.

Using Multiple Generative Learning Strategies

Although available research has mainly examined the effects of generative strategies in isolation, there are reasons to believe that using two generative strategies in a learning sequence would be even more effective than repeatedly using a single strategy. Previous research on using different generative learning strategies has shown that pairing strategies that serve unique cognitive processes may further enhance deep learning (Fiorella & Kuhlmann, 2020). For instance, self-explanation promotes knowledge organization and integration, which fosters comprehension and transfer (Bisra et al., 2018; Fiorella & Mayer, 2016). Whereas retrieval practice supports the process of retrieving information and mainly promotes retention (Roediger & Butler, 2011). Using self-explanation and retrieval practice in a learning sequence might allow learners to experience the direct and indirect benefits of both these strategies. In that case, retrieval practice might consolidate the elaborated mental models that were formed during self-explanation and promote retention of this knowledge over a longer period of time (Rittle-Johnson & Loehr, 2017).

So far, the limited research on using multiple generative learning strategies has mainly focused on using strategies simultaneously (e.g., Fiorella & Kuhlmann, 2020; Larsen et al., 2013; Roelle & Nückles, 2019). For example, a study by Larsen et al. (2013) examined testing and self-explaining among 47 university students. They found that participants in the ‘testing with self-explaining’ condition showed better learning outcomes during the initial study phase than participants in the ‘testing without explanations’ condition and the ‘restudy with and without self-explaining’ conditions.

Sequencing Multiple Generative Learning Strategies

Nevertheless, the efficacy of using multiple generative strategies rather than one might depend on the order in which the strategies are used. Firstly, to generate correct self-explanations learners should possess some prior knowledge (Renkl, 2014). According to elaborative retrieval theory, retrieval practice activates related knowledge in memory and encodes this along with the targeted information, which creates multiple memory traces to access this information in the future (Carpenter, 2009). Therefore, a retrieval practice (RP) – self-explanation (SE) sequence between study phases would allow learners to access and expand relevant prior knowledge through retrieval, before engaging in self-explanation. A self-explanation (SE) – retrieval practice (RP) sequence would likely hinder the generating of self-explanations by a lack of prior knowledge, which would in turn lead to low-quality mental representations of the learning material (Hiller et al., 2020). Consolidating these deficient mental representations in a subsequent retrieval practice activity would offer few benefits in terms of learning outcomes.

Contrarily, self-explaining might especially benefit learners early in the study phase when they lack high-quality mental representations, because generative processing increases the coherence of learners' mental representations and their integration with prior knowledge (Roelle & Nückles, 2019). At the end of the study phase, learners likely possess high-quality mental representations. Therefore, in a RP-SE sequence, self-explanation might be redundant since this strategy mainly focuses on the formation of high-quality mental representations and those are likely present at the end of the study phase (Roelle & Nückles, 2019). A SE-RP sequence would allow learners to first form high-quality mental representations through generative processing, presumably also because this enables learners to recognize potential knowledge gaps which they can repair during the remaining study phase (Lachner et al., 2020). Furthermore, it might be possible that learners are more able to detect knowledge gaps after self-explaining than after retrieval practice, since learners receive relatively sparse feedback about knowledge gaps in a free-recall task (Endres et al., 2020). Therefore, in a SE-RP sequence, retrieval practice might be more beneficial at the end of the study phase, because a subsequent retrieval practice activity would consolidate the high-quality mental representations that learners formed as a result of self-explaining (Roelle & Nückles, 2019).

The Present Study

Given the need for additional research on the effects of using multiple generative learning strategies in a learning sequence, this study examined the effects of self-explanation and retrieval practice in secondary vocational education students while learning from instructional videos. Both generative learning strategies have shown to be suitable and effective for students in this age group (Brod, 2020). Additionally, existing studies mainly examined the use of multiple strategies simultaneously. Since the use of multiple strategies simultaneously can pose too much cognitive load (cf. cognitive load theory) and different strategies are beneficial for distinct phases of learning (Fiorella & Mayer, 2016), this study focuses on the use of multiple strategies consecutively. The primary research question was whether engaging in multiple generative learning strategies rather than a single strategy would increase learning from instructional videos. Effects on retention, comprehension, and transfer were measured on a delayed posttest to investigate the effects on learning outcomes. Previous research has shown that learners who employ multiple generative learning strategies produce better learning outcomes, because different strategies serve unique cognitive processes that enhance learning (Fiorella & Kuhlmann, 2020; Koh et al., 2018; Rittle-Johnson & Loehr, 2017). Contrarily, other studies did not find beneficial effects of using multiple strategies rather than a single learning strategy (Roelle & Nückles, 2019; Waldeyer et al., 2020). Additionally, since every generative learning strategy comes with a cost, for example more time or cognitive demands (Bisra et al, 2018; Rittle-Johnson & Loehr, 2017), use of multiple strategies might provide learners with double the costs which might negatively impact their learning outcomes. However, it is also possible that a variance in learning strategies increases learners' academic achievement and keeps them motivated during learning (Magen-Nagar & Cohen, 2017). Therefore, no hypothesis on the use of multiple strategies rather than a single strategy was formed. Additionally, this study examined whether the sequence of self-explanation and retrieval practice (SE-RP vs. RP-SE) affected learning. Since different theories contradict one another and previous research is limited, this was approached as an open question and no hypothesis was formed. Furthermore, based on the cognitive processes involved, one might expect that self-explaining (SE-SE) improves higher order learning such as comprehension and transfer, whereas retrieval practice (RP-RP) mainly promotes lower order learning such as retention.

Furthermore, effects on perceived mental effort were measured because previous research has shown that perceived mental effort influences students’ learning strategy decisions (Hui et al., 2022). Additionally, effects on metacomprehension accuracy were explored to obtain more information on students’ metacognitive processes (Gutierrez de Blume, 2022). Lastly, learning enjoyment and self-efficacy were also measured, since these are important components of motivation and therefore they influence students’ learning strategy decisions and willingness to use a certain strategy (Chan et al., 2012; Sadi & Uyar, 2013). Since the analyses of the abovementioned variables were explorative, no hypotheses were formed.

Method

Participants and Design

Participants were Dutch secondary vocational education students. In the experiment a total of 155 students (83 male, 72 female) with a mean age of 20.12 (*SD* = 4.07) participated. Of those 155 students, 95 students (50 male, 45 female) with a mean age of 19.68 (*SD* = 3.97) completed the delayed posttest. G*Power analysis revealed that in order to detect a medium effect size with 80% power ($\alpha = .05$), a minimum of 128 total participants was required (Faul et al., 2007). A medium effect size was used for the power analysis since this effect size is the threshold of educational relevance (Hattie, 2009). This study employed a between-subjects design with four conditions in order to compare the effects of using multiple strategies and sequencing on retention, comprehension, and transfer across the different groups. Participants were randomly assigned to either a RP-RP, SE-SE, RP-SE, or SE-RP condition, see Table 1. Participants had little to no prior knowledge on the subject of the video, since this was not covered in their curriculum. Participation in the experiment was voluntary and no compensation was provided.

Table 1

Distribution of Participants Among Conditions

		N	
		Experiment	Delayed posttest
Condition	RP-RP	40	25
	SE-SE	38	22
	RP-SE	38	23
	SE-RP	39	25
Total		155	95

Materials

The materials for this study consisted of a video on the pollination of plants, a retrieval practice task, a self-explanation task, five questionnaires to measure several constructs, and a delayed posttest. All materials were presented via Qualtrics in Dutch, as this was the native language of most participants.

Prior Knowledge Questionnaire

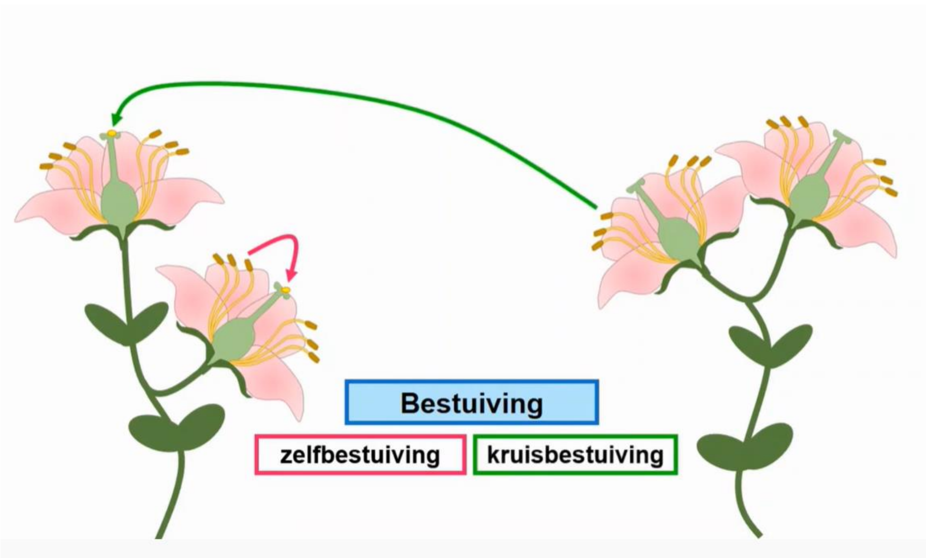
A questionnaire rather than a pretest was used to measure prior knowledge to avoid a testing effect, where the act of taking a test enhances later performance (Roediger & Butler, 2011). The prior knowledge questionnaire asked participants to rate their prior knowledge on the pollination of plants on a 5-point Likert scale, with 1 representing ‘very low’ and 5 representing ‘very high’. Additionally, the questionnaire included five statements relating to knowledge about pollination, where students had to mark the statements that applied to them. Examples of these statements include: “I know exactly what the pollination of plants is.” and “I know exactly which processes are required for the pollination of plants.”. Cronbach’s alpha for the prior knowledge questionnaire was $\alpha = .76$, which is considered acceptable reliability (Tavakol & Dennick, 2011).

Instructional Video

Participants watched an instructional video about the pollination of plants (see Figure 1), with a duration of 3 minutes and 51 seconds. The video showed the reproductive organs of plants, different methods plants use to spread pollen, and the two types of pollination. The video was completely narrated by a teacher with a detailed explanation of the process of pollination.

Figure 1

Screenshot from the Instructional Video about Pollination



Learning Tasks

The retrieval practice task consisted of a free-recall prompt where participants were instructed to write down everything they remembered about the contents of the video. The self-explanation task included a prompt that directed participants to generate a written explanation based on what they learned from the video about pollination.

Posttest

The delayed posttest consisted of 19 questions in total, of which 13 were open-ended questions and six multiple choice questions. The posttest intended to measure retention, comprehension, and transfer. A total of seven items measured retention, such as: ‘Name the two methods plants use to disperse pollen'. Furthermore, five items measured comprehension with questions such as: 'Explain why wind-pollinated flowers produce more pollen than insect-pollinated flowers'. Lastly, seven items measured transfer by requiring participants to apply the knowledge they gained from the video to a new situation. For example, participants were asked to explain why a single fruit tree in a garden did not bear any fruit. This question was not directly discussed in the video about pollination, but participants could use the knowledge they gained from the video to answer this question correctly. While Cronbach's alpha for the total posttest was $\alpha = .80$, which is considered acceptable, reliability for retention ($\alpha = .65$), comprehension ($\alpha = .61$), and transfer ($\alpha = .54$) was questionable (Tavakol & Dennick, 2011). Moreover, removal of items for these variables would not lead to higher reliability.

Additionally, a Confirmatory Factor Analysis (CFA) was conducted to determine construct validity. Results showed that the CFA model with the factors retention, comprehension, and transfer did not fit well with the data. The Chi-square goodness-of-fit test was significant ($p < .001$) and the fit indices were insufficient (CFI =0.795; TLI = 0.765; RMSEA = 0.069). Thus, Exploratory Factor Analysis (EFA) was conducted to identify the underlying structure of the measured variables. The model with three factors was non-significant, which means it fitted the data ($p = .469$). However, inspection of factor loadings did not show any logical theoretical structure of the items. Therefore, interpretation of these three factors was not possible and the original intended constructs were used for the remainder of this study.

Perceived Mental Effort

Perceived mental effort was measured by asking participants to indicate how much mental effort they had invested after watching each video and learning task on a 9-point Likert scale ranging from 1 'very, very low effort' to 9 'very, very high effort'. This scale was developed by Paas (1992) and has shown consistent reliability, with reliability coefficients ranging from ≥ .82 – .93 (Van Gog & Paas, 2008).

Metacomprehension Accuracy

To measure metacomprehension accuracy, participants were asked to make judgements of learning about their expected performance on the posttest (for example, see Lachner et al., 2020). Participants were asked to estimate how many questions they would answer correctly on the posttest, ranging from 0 to 30. Metacomprehension accuracy was operationalized in terms of bias.

Learning Enjoyment

Learning enjoyment was measured by asking participants to indicate how much enjoyment they experienced during the learning tasks on a 9-point Likert scale ranging from 1 'very, very low enjoyment to 9 'very, very high enjoyment' (Hoogerheide et al., 2019).

Self-Efficacy

Self-efficacy was also measured on a 9-point Likert scale, asking participants to rate their confidence about their knowledge on the pollination of plants (cf. Bandura, 2006; see also Hoogerheide et al., 2016), on a scale ranging from 1 'very, very low confidence' to 9 'very, very high confidence'.

Procedure

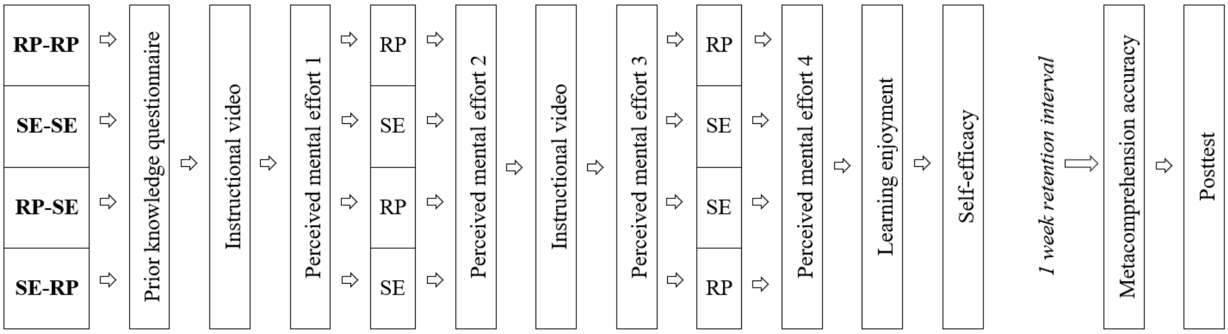
First, a pilot study of the experiment was carried out with 17 students that represented the target group. After conducting the pilot study, some changes were made to improve the design of the study. For example, the instructions and questioning were further clarified. Additionally, because some students experienced technical difficulties with the automatic timer in Qualtrics, it was decided that the researcher would keep track of time during the learning tasks.

Subsequently, the experiment was conducted at participants' school using laptops. See Figure 2 for an overview of the study's procedure. To protect participants' identities, all participants received a de-identified participant number that was used throughout the experiment. First, the experimenter provided a plenary introduction to the study and instructions to access Qualtrics. When the participants accessed Qualtrics, they received an informative letter and the opportunity to provide informed consent. Next, participants completed the prior knowledge questionnaire and watched the video on pollination for the first time. Participants were instructed to watch the video alone and to use headphones to minimize distractions. Subsequently, participants had 4 minutes to complete either a retrieval practice task or a self-explanation task. During the learning tasks, participants did not have access to the learning materials. For the self-explaining task, participants were instructed to explain everything they could remember about the contents of the video as if they were explaining it to a fellow student who had no knowledge on plant pollination. For the retrieval practice task, participants were instructed to write down everything they could remember about the contents of the video. Participants in both conditions were instructed to write up their descriptions as complete as possible. After completing the learning task, participants watched the video for the second time. Participants then performed the final retrieval practice or self-explanation task for 4 minutes. During the whole experiment, participants had to indicate their perceived mental effort after each video and learning task. Additionally, after completing the final learning task participants reported how much enjoyment they experienced while working on the learning tasks and how confident they felt about their knowledge on pollination. The total duration of the experiment was approximately 30 minutes.

After one week participants completed a posttest (10 min) to measure their retention, comprehension, and transfer. Beforehand, participants provided a judgement of learning. For the posttest, participants were instructed to answer the questions independently and not to use any other resources than their own memory. Due to mandatory school holidays and quarantines some students (n = 10) had to complete the posttest online and some students (n = 11) completed the posttest in more or fewer days than the intended one-week delay between the experiment and posttest.

Figure 2

Overview of the Study Procedure for all Conditions.



Data Analysis

The total score for self-reported prior knowledge ranged from 1 to 10 and was computed by adding the score on the Likert-scale (range 1 to 5) and one point for each marked statement about participants' prior knowledge (range 0 to 5). To compute total scores for retention, comprehension, and transfer, participants received one point for each correctly answered question or part of a question. Scores for retention, comprehension, and transfer ranged from 0 to 10 and thus the total posttest score ranged from 0 to 30. Two raters scored approximately 20% of the posttests using an answer model. Inter-rater reliability was excellent for retention, *ICC* = 0.99; comprehension, *ICC* = 0.93; and transfer, *ICC* = 0.99 (Koo & Li, 2016). Therefore, only one rater coded the remaining posttests. Additionally, scores for metacomprehension accuracy bias were computed. Due to an error in the answer option of the question on metacomprehension accuracy, some participants did not report a correct value. These answers (n = 10) could not be used and were treated as missing values.

Metacomprehension accuracy bias was computed for the remaining sample (*n* = 85) by calculating the difference between the estimated number of correct questions and the actual number of correct questions. Positive values indicated overestimation of performance, negative values indicated underestimation of performance, and values of zero showed accurate judgement of learning.

Results

All analyses were conducted with an alpha level of .05. For the analyses, two-way ANOVAs, one-way and two-way MANOVAs, and Kruskal-Wallis tests were used. For analyzing multiple dependent variables, MANOVAs were used instead of ANOVAs to avoid the increased risk of Type 1 error that can occur when analyzing related dependent variables (Meyers et al., 2016). Effect sizes were measured by partial η^2, interpreting values <.06 as small effects, values ranging from 0.06 to 0.14 as medium effects, and values > 0.14 as large effects (Cohen, 1988). Furthermore, all assumptions were checked following the guidelines of Field (2018) and any violations of assumptions were reported.

Data for the preliminary analyses, retention, comprehension, transfer, total posttest, perceived mental effort, self-efficacy, and learning enjoyment were not normally distributed as assessed by conducting a Shapiro-Wilk test ($p < .05$). Therefore, skewness and kurtosis values were computed to determine if these violations of normality were cause for concern. Values for skewness and kurtosis should lie between -2 and +2 to be considered acceptable in order to prove normal distribution (George & Mallery, 2010). All variables showed acceptable skewness and kurtosis values. Therefore, since ANOVAs and two-way MANOVAs are also considered to be fairly robust to deviations from normality (Bray & Maxwell, 1985; Maxwell & Delaney, 2004; Weinfurt, 1995) and skewness and kurtosis values were generally acceptable, the analyses were conducted anyway.

Furthermore, data for the preliminary analyses, comprehension, transfer, total posttest, perceived mental effort, and learning enjoyment showed several univariate outliers. Therefore, raw data was examined for possible errors or inaccuracies, however no clear errors or causes were found. Additionally, *z*-scores were computed to determine if these outliers could be cause for concern. According to Field (2018), values of *z*-scores should lie between -3.29 and 3.29. Since none of the cases showed extreme *z*-scores on the dependent variables, outliers were not removed for further analyses. Additionally, data for perceived mental effort showed two multivariate outliers as assessed by computing Mahalanobis distance. Because scores for these two multivariate outliers were above the critical Chi-square value of 16.27 ($\alpha = .001$), they were considered multivariate outliers (Tabachnick & Fidell, 2014). However, raw data did not show any errors or inaccuracies, so therefore the analysis for perceived mental effort was conducted with and without these multivariate outliers.

Lastly, data for prior knowledge, retention, comprehension, transfer, and total posttest score violated the assumption of homogeneity of variances as assessed by conducting Levene's Test ($p < .05$). However, when group sizes are approximately the same size, which is the case, F tests are relatively insensitive to the violation of this assumption (Chen & Zhu, 2001). Hence, the analyses were carried out anyway.

Descriptive statistics of the study are presented in Table 2. The sample sizes that were used to compute the mean scores differed between variables, since some variables were measured during the experiment while other variables were measured during the delayed posttest. The table shows that descriptively, for the sample as a whole, test performance for retention, comprehension, and transfer did not approach a floor or ceiling effect and could be classified as moderate.

Table 2

Descriptive Statistics for Study Variables

	RP-RP	SE-SE	RP-SE	SE-RP
Retention[a] *(range 0 – 10)*	4.20 (2.69)	4.14 (2.27)	4.52 (2.87)	4.92 (2.57)
Comprehension[a] *(range 0 – 10)*	3.64 (2.31)	3.32 (1.89)	3.43 (2.73)	4.00 (1.61)
Transfer[a] *(range 0 – 10)*	5.08 (1.58)	5.32 (1.09)	5.26 (1.82)	5.92 (1.19)
Total score posttest[a] *(range 0 – 30)*	12.92 (5.69)	12.77 (4.13)	13.22 (6.45)	14.84 (3.88)
Metacomprehension accuracy prediction[b] *(range 0 – 30)*	15.92 (5.20)	17.25 (7.50)	15.81 (5.52)	17.00 (3.92)
Metacomprehension accuracy bias[b] *(range -30 – 30)*	-3.00 (6.84)	-4.25 (6.05)	-2.67 (7.28)	-1.47 (4.33)
Mental effort after first video[c] *(range 1 – 9)*	4.25 (1.84)	3.87 (2.02)	3.42 (2.06)	3.49 (1.90)
Mental effort after first learning task[c] *(range 1 – 9)*	4.48 (1.97)	4.76 (2.02)	4.63 (2.19)	4.13 (2.15)
Mental effort after second video[c] *(range 1 – 9)*	4.18 (1.84)	3.79 (2.16)	3.63 (1.84)	4.00 (1.95)
Mental effort after second learning task[c] *(range 1 – 9)*	4.37 (1.68)	4.47 (1.98)	4.63 (1.94)	4.62 (1.60)
Self-efficacy[c] *(range 1 – 9)*	4.98 (1.64)	5.29 (1.84)	4.82 (1.89)	5.28 (2.00)
Learning enjoyment[c] *(range 1 – 9)*	4.97 (1.56)	4.37 (1.87)	4.97 (2.11)	5.08 (1.77)

[a] n = 95. [b] n = 85. [c] n = 155.

Preliminary Analyses

Several preliminary analyses were conducted to test whether the random assignment led to comparable conditions. First, differences between experimental conditions concerning prior knowledge were checked using a two-way ANOVA with the first learning task (RP vs. SE) and the second learning task (RP vs. SE) as between-subject factors. The analysis showed no significant main effects (First learning task: $F(1, 91) = 1.29$, $p = .259$, $\eta^2_p = .014$; Second learning task: $F(1, 91) = 1.14$, $p = .288$, $\eta^2_p = .012$) or interaction effect ($F < 1$). This indicated that there were no significant differences in prior knowledge between conditions before starting with the experiment.

Additionally, since some participants completed the posttest online, a one-way MANOVA was conducted with retention, comprehension, and transfer as the dependent variables to determine if there were any differences between participants who took the posttest online and participants who took the posttest at school. Results showed no significant differences between these participants, $F(3, 91) = 0.02$, $p = .997$, Wilks' $\Lambda = .999$, $\eta^2_p = .001$. This means that participants who completed the posttest online were comparable with participants who took the posttest at school and any potential effects of condition cannot be attributed to the way that participants took the test.

Lastly, a one-way MANOVA was conducted to compare participants who had more or fewer days between the experiment and the posttest than other participants who had exactly one week between the experiment and the posttest. Results of the analysis showed that the differences between these participants on retention, comprehension, and transfer were non-significant, $F(12, 233) = 0.82$, $p = .628$, Wilks' $\Lambda = .896$, $\eta^2_p = .036$. This indicated that the time between the experiment and the posttest did not affect participants' achievement on the posttest.

Learning Outcomes

The effects of using multiple strategies and sequencing on retention, comprehension, and transfer were analyzed with a two-way MANOVA with first task (RP vs. SE) and second task (RP vs. SE) as between-subject factors. The interaction effect between first task and second task on the combined dependent variables was not statistically significant, $F(3, 89) = 0.75$, $p = .523$, Wilks' $\Lambda = .975$, $\eta^2_p = .025$. Furthermore, the main effects of the first task and the second task were not statistically significant (First learning task: $F(3, 89) = 0.88$, $p = .454$, Wilks' $\Lambda = .971$, $\eta^2_p = .029$; Second learning task: $F(3, 89) = 0.35$, $p = .791$, Wilks' $\Lambda = .988$, $\eta^2_p = .012$). These results indicated that the amount of generative learning strategies (one vs.

two) and the order in which different strategies were used did not affect students' retention, comprehension, and transfer.

Furthermore, removal of the previously mentioned univariate outliers also did not lead to a significant interaction effect of first and second learning task, $F(3, 87) = 1.04$, $p = .378$, Wilks' $\Lambda = .965$, $\eta^2_p = .035$. Likewise, the main effect for first learning task was not significant without the univariate outliers, $F(3, 87) = 0.91$, $p = .440$, Wilks' $\Lambda = .970$, $\eta^2_p = .030$, nor for the second learning task, $F(3, 87) = 0.34$, $p = .798$, Wilks' $\Lambda = .988$, $\eta^2_p = .012$.

Since quality checks showed low reliability for the separate constructs of the posttest, whereas the total posttest score had acceptable reliability, it was explored whether effects of using multiple strategies or sequencing on the total posttest score could be found. A two-way ANOVA was conducted with the first and second learning task as between-subject factors and the total posttest score as the dependent variable. However, no significant interaction effect, $F(1, 91) = 1.25$, $p = .267$, $\eta^2_p = .014$, or main effects of first and second learning task were found (First learning task: $F(1, 91) = 0.49$, $p = .487$, $\eta^2_p = .005$; Second learning task: $F(1, 91) = 0.70$, $p = .405$, $\eta^2_p = .008$). This indicated that, in line with the previous analysis, the amount of generative learning strategies or the sequence in which these strategies were performed, did not affect students' learning outcomes.

Lastly, considering that several assumptions for the MANOVA were violated, a non-parametric test was conducted to check for the robustness of the findings. Since SPSS does not offer non-parametric alternatives for a MANOVA or two-way ANOVA, three separate Kruskal-Wallis tests were conducted to determine if there were differences in retention, comprehension, and transfer scores between conditions. Results showed that there were no statistically significant differences between conditions in terms of retention, $H(3) = 1.69$, $p = .639$; comprehension, $H(3) = 1.88$, $p = .597$; and transfer, $H(3) = 5.36$, $p = .148$.

To conclude, the lack of significant findings indicated that the groups did not differ on retention, comprehension, and transfer or the total posttest score. These results showed that the use of one or two generative learning strategies did not affect learning outcomes. Similarly, the sequence in which self-explanation and retrieval practice were performed did not affect students' learning.

Explorative Analyses

Several explorative analyses were performed to determine if there were any effects of using multiple strategies or sequencing on perceived mental effort, self-efficacy, metacomprehension accuracy bias, and learning enjoyment. A two-way MANOVA was conducted to understand the effects of first learning task (RP vs. SE) and second learning task (RP vs. SE) on perceived mental effort scores. Perceived mental effort scores after the first and second learning task were used as dependent variables to see if participants experienced different levels of effort while using different strategies. Additionally, perceived mental effort after watching the video for the second time was used as a dependent variable to see if there were any carry-over effects from the strategies used prior to watching the video. Perceived mental effort after watching the video for the first time was not used as a dependent variable, since participants did not differ in prior knowledge and no manipulations were used in that stage of the experiment. Results of the analysis showed no significant interaction effect of first and second learning task ($F < 1$). Moreover, no significant effects of first learning task ($F < 1$), or second learning task, $F(3, 149) = 2.16$, $p = .095$, Wilks' $\Lambda = .958$, $\eta^2_p = .042$, were found. Removal of the previously mentioned multivariate outliers also did not lead to any significant findings. These results indicated that the groups did not experience different levels of perceived mental effort during the learning tasks or while watching the video for the second time.

Furthermore, to explore if there were any differences between conditions regarding self-efficacy, a two-way ANOVA with first and second learning task as between-subject factors was conducted. Yet, no significant main effects of first learning task, $F(1, 151) = 1.73$, $p = .190$, $\eta^2_p = .011$, second learning task ($F < 1$), or interaction effect ($F < 1$) were found. This suggested that participants across conditions felt confident in their knowledge of the subject of the video in a comparable manner.

To measure the effects of the first and second learning task on students' metacomprehension accuracy bias, a two-way ANOVA was conducted with first and second learning task as between-subject factors. Results showed no significant main effects of first and second learning task ($F < 1$), or interaction effect, $F(1, 81) = 1.28$, $p = .261$, $\eta^2_p = .016$. This indicated that the use of one or two generative strategies and the sequence of these strategies did not affect the accuracy of students' judgement of learning.

Lastly, a two-way ANOVA was conducted to examine if the conditions differed regarding learning enjoyment. However, results showed no significant main effect of first learning task ($F < 1$), second learning task, $F(1, 151) = 1.45$, $p = .230$, $\eta^2_p = .010$, or interaction effect, $F(1, 151) = 1.44$, $p = .232$, $\eta^2_p = .009$. This means that participants across conditions experienced comparable levels of learning enjoyment.

Discussion

The aim of the present study was to examine the effects of self-explanation and retrieval practice while learning from instructional videos in secondary vocational education students. More specifically, the purpose of this study was to examine the effects of using multiple rather than a single generative learning strategy and the sequence in which the two strategies were used, on students' retention, comprehension, and transfer. Previous studies called for additional research on the effects of sequencing different generative learning strategies (Fiorella & Mayer, 2016; Lachner et al., 2021). Therefore, the present study aimed to fill this gap in order to extent generative learning theory and provide practical implications on the use of multiple generative learning strategies.

Findings of this study showed no significant effects of using multiple generative learning strategies or the sequence of these strategies on retention, comprehension, or transfer as well as on the total posttest score. This means there was no difference in learning outcomes between students who either used one or two learning strategies or between students who performed self-explanation and retrieval practice in a different sequence. Similarly, no significant effects of using multiple generative learning strategies or the sequence of retrieval practice and self-explanation were found on self-efficacy, perceived mental effort, metacomprehension accuracy, or learning enjoyment.

The finding that using two different strategies was not more effective for learning than using a single strategy is in line with prior studies by Roelle and Nuckles (2019; retrieval practice and explaining/giving examples) and Waldeyer et al. (2020; retrieval practice and explaining/giving examples). Yet, contrasts other studies by Fiorella and Kuhlmann (2020; drawing and teaching), Koh et al. (2018; retrieval practice and teaching), and Larsen et al. (2013; testing and self-explaining) that did find beneficial effects of using multiple generative learning strategies on learning outcomes. Additionally, the absence of an effect of sequencing learning strategies is consistent with the study by Fiorella et al. (2021), who found no effect of the sequence of explaining and drawing activities on learning outcomes. However, a recent study by Roelle et al. (2022)[1] did find an effect of sequencing retrieval practice and example generation. More specifically, they found that students who performed retrieval practice before giving examples showed higher learning gains, experienced lower cognitive load, and needed less time-on-task. For a comprehensive overview of previous research on using multiple strategies or sequencing of different strategies, see Appendix A.

[1] The study by Roelle et al. (2022) was published when the present study was already conducted. Therefore, it was not mentioned in the Introduction section.

There are several possible explanations on why no significant effects of using multiple strategies or sequencing on learning outcomes were found. Firstly, one possible explanation lies in the extent to which different strategies differ in terms of cognitive processes that are triggered by a strategy. The present study used two learning strategies with different cognitive functions, with retrieval practice serving a consolidation function and self-explaining mainly supporting the construction of mental models (Roediger & Butler, 2011; Wylie & Chi, 2014). However, certain choices in the design of the experiment might have weakened the distinction between the two different learning tasks. More specifically, this study made use of free recall tasks as opposed to targeted retrieval tasks. Previous research showed that free recall leads to increased retention of broader non-targeted information, whereas targeted retrieval supports retention of targeted information from the learning contents (Endres et al., 2020). Furthermore, participants did not have access to the learning materials while performing the learning tasks. This is especially important for the self-explanation task, since the absence of the learning materials during the task automatically required participants to retrieve the contents of the video from their memory to generate their explanations. Therefore, the more elaborative nature of the free-recall task and the required retrieval practice for the self-explanation task might have weakened the distinction between the two different learning tasks, which in turn made the conditions more similar. Using two strategies that are very similar might be comparable to repeatedly using a single strategy, which might explain the absence of a significant effect of using multiple strategies. Likewise, sequencing similar strategies might be comparable to repeatedly using the same strategy, because the same cognitive processes are stimulated throughout the entire learning sequence. In that case, the transition to a different strategy in the learning sequence might be indistinctive. This distinction between strategies would also explain why several previous studies (e.g. Fiorella & Kuhlmann, 2020; Larsen et al., 2013; Roelle et al., 2022) did find an effect of using multiple strategies or sequencing, while others did not (see Appendix A). For example, Fiorella et al. (2021) did not find an effect of sequencing explaining and drawing tasks, which support the same cognitive processes, namely that of knowledge organization and integration. Whereas Roelle et al. (2022) did find a significant effect of sequencing on learning outcomes, and they used open-book example generation tasks, which support knowledge organization and integration, and more targeted recall cues, that serve a consolidation function. Therefore, the distinction between the cognitive processes induced by different strategies might serve as a potential boundary condition that has to be met, before the use of multiple strategies or sequencing has an effect on learning outcomes.

Another possible explanation for the lack of significant findings, is that participants might have not succeeded in producing reasonable-quality explanations or successful retrieval attempts. Since participants' answers were not scored, the quality of self-explanations and retrieval attempts remains unknown. However, the effectiveness of self-explaining depends on the quality of the explanations, yet students might have needed feedback or training to generate reasonable-quality explanations (e.g. Berthold et al., 2009; Lachner & Neuburg, 2019; Rittle-Johnson & Loehr, 2017). Similarly, successful retrieval is critical for learning from retrieval practice, considering that the beneficial effects of retrieval practice decrease or disappear altogether when students do not retrieve the correct information (Roediger & Butler, 2011). Participants might have needed feedback on their retrieval attempts before they could successfully retrieve the correct information, since providing feedback on retrieval attempts not only enhances student performance, but also supports future successful retrieval attempts (Karpicke & Grimaldi, 2012; Roediger & Butler, 2011; Rowland, 2014). Therefore, the absence of feedback on participants' answers to the learning tasks or the lack of training on how to use these learning strategies might have led to low-quality explanations and failed retrieval attempts, which decreased the beneficial effects of these learning strategies on students' learning across all conditions.

Furthermore, the limited time between the study phase and learning tasks in this experiment might explain why there was no significant effect of retrieval practice on retention. According to retrieval practice theory, when the delay between the initial study phase and the retrieval phase is short, the temporal context remains comparable. This in turn limits the formation of distinctive context cues that can be used to retrieve the learning materials in the future (Karpicke et al., 2014). Therefore, the limited time between the study phase and learning tasks might have hindered the formation of retrieval cues. A lack of retrieval cues might have hampered the beneficial effects of retrieval practice on retention. Considering that retrieval practice was also necessary for students in the self-explanation condition, problems with retrieving relevant information prior to self-explaining might also explain why there was no significant effect of self-explaining on comprehension and transfer. Either because participants failed to retrieve the required information from memory or because this retrieval hurdle took a substantial amount of time (Hiller et al., 2020).

Besides previously mentioned limitations in the design of the study (i.e., weak distinction between learning tasks, lack of scoring and feedback on explanations and retrieval attempts), there are methodological limitations to this study that might explain why no effects of using multiple strategies or sequencing on learning outcomes were found. First, the measurement instruments for retention ($\alpha = .65$), comprehension ($\alpha = .61$), and transfer ($\alpha = .54$) showed low reliability (Field, 2018). Low reliability is problematic, considering that it generates scores that do not precisely reflect students' true performance (Furr, 2021). Furthermore, while the reliability of the total posttest was acceptable ($\alpha = .80$), factor analysis showed problems with its validity. This suggests that the posttest was not suitable to measure the constructs of retention, comprehension, and transfer. Hence, if the posttest did not measure the intended variables and the reliability of the measurements was also low, it seems sensible that no effects on learning outcomes were found.

As for the practical implications, these findings suggest that teachers might not have to make use of two strategies rather than one or sequencing of strategies to increase student performance. However, the added value of generative learning strategies is already well-established in literature (e.g. Fiorella & Mayer, 2016). Therefore, educators should encourage students to use generative learning strategies to enhance their learning. Furthermore, educators should especially incorporate generative learning strategies with instructional videos, because this transforms video learning into an active learning experience, which has proven to foster learning (e.g., Fiorella et al., 2020; Kleiman et al., 2019; Mayer et al., 2020). Regarding theoretical implications, these findings suggest there might be several boundary conditions that have to be met before the use of multiple strategies or sequencing of strategies affects learning. Future research should use generative learning tasks that clearly differ in underlying cognitive processes and take the quality of students' answers on learning tasks into account by providing feedback or training on how to use a certain strategy. Furthermore, it would be interesting to investigate whether or not the time between the study phase and learning tasks affects the degree to which multiple strategies or sequencing are beneficial. More research is necessary to determine if these factors are indeed boundary conditions and before we can conclude that this is truly an addition to the theory on generative learning.

In sum, this study did not provide evidence for the beneficial effects of the use of two rather than one generative learning strategies or the sequencing of retrieval practice and self-explaining on students' learning outcomes. However, findings of this study should be interpreted with caution due to several limitations in the design of the experiment and low reliability and validity of certain measurement instruments. Still, this study adhered to the call for further research on sequencing different generative strategies. Additionally, recent research did show promising results regarding sequencing generative learning strategies. Therefore, future research is necessary to clarify possible boundary conditions and to examine the potential benefits of using multiple generative learning strategies and their sequence on learning.

References

Bae, C. L., Therriault, D. J., & Redifer, J. L. (2019). Investigating the testing effect: Retrieval as a characteristic of effective study strategies. *Learning and Instruction, 60*, 206–214. https://doi.org/10.1016/j.learninstruc.2017.12.008

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

Berthold, K., Eysink, T. H. S., & Renkl, A. (2009). Assisting self-explanation prompts are more effective than open prompts when learning with multiple representations. *Instructional Science*, *37*(4), 345–363. https://doi.org/10.1007/s11251-008-9051-z

Bisra, K., Liu, Q., Nesbit, J. C., Salimi, F., & Winne, P. H. (2018). Inducing self-explanation: A meta-analysis. *Educational Psychology Review*, *30*(3), 703–725. https://doi.org/10.1007/s10648-018-9434-x

Bray, J. H., & Maxwell, S. E. (1985). *Multivariate analysis of variance.* Sage. https://doi.org/10.4135/9781412985222

Brod, G. (2020). Generative learning: Which strategies for what age? *Educational Psychology Review*, *33*(4), 1295–1318. https://doi.org/10.1007/s10648-020-09571-9

Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology, 36*(5), 1118 –1133. https://doi.org/10.1037/a0019902

Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology*, *35*(6), 1563–1569. https://doi.org/10.1037/a0017021

Chan, K. W., Wong, K. Y. A., & Lo, E. S. C. (2012). Relational analysis of intrinsic motivation, achievement goals, learning strategies and academic achievement for

Hong Kong secondary students. *The Asia-Pacific Education Researcher, 21*(2), 230-243.

Chen, A., & Zhu, W. (2001). Revisiting the assumptions for inferential statistical analyses: A conceptual guide. *Quest, 53*(4), 418-439. https://doi.org/10.1080/00336297.2001.10491756

Chi, M. T. H. (2000). Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. In R. Glaser (Ed.), *Advances in Instructional Psychology* (pp. 161–238). Lawrence Erlbaum Associates.

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

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

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, *14*(1), 4-58. https://doi.org/10.1177/1529100612453266

Endres, T., Kranzdorf, L., Schneider, V., & Renkl, A. (2020). It matters how to recall: Task differences in retrieval practice. *Instructional Science, 48*(6), 699-728. https://doi.org/10.1007/s11251-020-09526-1

Endres, T., & Renkl, A. (2015). Mechanisms behind the testing effect: An empirical investigation of retrieval practice in meaningful learning. *Frontiers in Psychology*, *6*, 1-6. https://doi.org/10.3389/fpsyg.2015.01054

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191. https://doi.org/10.3758/BF03193146

Field, A. (2018). *Discovering statistics using IBM SPSS statistics.* Sage.

Fiorella, L., & Kuhlmann, S. (2020). Creating drawings enhances learning by teaching. *Journal of Educational Psychology*, *112*(4), 811-822. https://doi.org/10.1037/edu0000392

Fiorella, L., & Mayer, R. E. (2015). *Learning as a generative activity.* Cambridge University Press. https://doi.org/10.1017/cbo9781107707085

Fiorella, L., & Mayer, R. E. (2016). Eight ways to promote generative learning. *Educational Psychology Review*, *28*(4), 717-741. https://doi.org/10.1007/s10648-015-9348-9

Fiorella, L., Pyres, M., & Hebert, R. (2021). Explaining and drawing activities for learning from multimedia: The role of sequencing and scaffolding. *Applied Cognitive Psychology*, *35*(6), 1574-1584. https://doi.org/10.1002/acp.3871

Fiorella, L., Stull, A. T., Kuhlmann, S., & Mayer, R. E. (2020). Fostering generative learning from video lessons: Benefits of instructor-generated drawings and learner-generated explanations. *Journal of Educational Psychology*, *112*(5), 895-906. https://doi.org/10.1037/edu0000408

Furr, R. M. (2021). *Psychometrics: An introduction.* SAGE publications.

George, D., & Mallery, M. (2010). *SPSS for Windows step by step: A simple guide and reference.* (10th ed.) Pearson. https://doi.org/10.4324/9781003205333

Gutierrez de Blume, A. P. (2022). Calibrating calibration: A meta-analysis of learning strategy instruction interventions to improve metacognitive monitoring accuracy. *Journal of Educational Psychology*, *114*(4), 681–700. https://doi.org/10.1037/edu0000674

Hattie, J. A. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement.* Routledge.

Hiller, S., Rumann, S., Berthold, K., & Roelle, J. (2020). Example-based learning: Should learners receive closed-book or open-book self-explanation prompts? *Instructional Science*, *48*(6), 623-649. https://doi.org/10.1007/s11251-020-09523-4

Hoogerheide, V., van Wermeskerken, M., Loyens, S. M., & van Gog, T. (2016). Learning from video modeling examples: Content kept equal, adults are more effective models than peers. *Learning and Instruction*, *44*, 22-30. https://doi.org/10.1016/j.learninstruc.2016.02.004

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*, 101-226. https://doi.org/10.1016/j.learninstruc.2019.101226

Hui, L., de Bruin, A. B., Donkers, J., & van Merriënboer, J. J. (2022). Why students do (or do not) choose retrieval practice: Their perceptions of mental effort during task performance matter. *Applied Cognitive Psychology, 36*(2), 433-444. https://doi.org/10.1002/acp.3933

Kang, S. H., McDermott, K. B., & Roediger III, H. L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology, 19*(4), 528-558. https://doi.org/10.1080/09541440601056620

Karpicke, J. D., & Grimaldi, P. J. (2012). Retrieval-based learning: A perspective for enhancing meaningful learning. *Educational Psychology Review*, *24*(3), 401-418. https://doi.org/10.1007/s10648-012-9202-2

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-based learning: An episodic context account. In B. Ross (Ed.), *The psychology of learning and motivation* (pp. 234–287). Elsevier. https://doi.org/10.1016/B978-0-12-800283-4.00007-1

Kleiman, A. M., Potter, J. F., Bechtel, A. J., Forkin, K. T., Dunn, L. K., Collins, S. R., Lyons, G., Nemergut, E. C., & Huffmyer, J. L. (2019). Generative retrieval results in positive academic emotions and long-term retention of cardiovascular anatomy using transthoracic echocardiography. *Advances in Physiology Education*, *43*(1), 47-54. https://doi.org/10.1152/advan.00047.2018

Kliegl, O., & Bäuml, K. H. T. (2021). When retrieval practice promotes new learning: The critical role of study material. *Journal of Memory and Language*, *120*, 1-13. https://doi.org/10.1016/j.jml.2021.104253

Koh, A. W. L., Lee, S. C., & Lim, S. W. H. (2018). The learning benefits of teaching: A retrieval practice hypothesis. *Applied Cognitive Psychology*, *32*(3), 401-410. https://doi.org/10.1002/acp.3410

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

Lachner, A., Backfisch, I., Hoogerheide, V., van Gog, T., & Renkl, A. (2020). Timing matters! Explaining between study phases enhances students' learning. *Journal of Educational Psychology*, *112*(4), 841-853. https://doi.org/10.1037/edu0000396

Lachner, A., Hoogerheide, V., van Gog, T., & Renkl, A. (2021). Learning-by-teaching without audience presence or interaction: When and why does it work? *Educational Psychology Review*, *34*(2), 1-33. https://doi.org/10.1007/s10648-021-09643-4

Lachner, A., Ly, K.-T., & Nuckles, M. (2018). Providing written or oral explanations? Differential effects of the modality of explaining on students' conceptual learning and transfer. *Journal of Experimental Education*, *86*(3), 344–361. https://doi.org/10.1080/00220973.2017.1363691

Lachner, A., & Neuburg, C. (2019). Learning by writing explanations: Computer-based feedback about the explanatory cohesion enhances students' transfer. *Instructional Science*, *47*(1), 19-37. https://doi.org/10.1007/s11251-018-9470-4

Latimier, A., Peyre, H., & Ramus, F. (2021). A meta-analytic review of the benefit of spacing out retrieval practice episodes on retention. *Educational Psychology Review*, *33*(3), 959-987. https://doi.org/10.1007/s10648-020-09572-8

Larsen, D. P., Butler, A. C., & Roediger III, H. L. (2013). Comparative effects of test-enhanced learning and self-explanation on long-term retention. *Medical Education, 47*(7), 674-682. https://doi.org/10.1111/medu.12141

Lawson, A. P., & Mayer, R. E. (2021). Benefits of writing an explanation during pauses in multimedia lessons. *Educational Psychology Review*, *33*(4), 1859-1885. https://doi.org/10.1007/s10648-021-09594-w

Magen-Nagar, N., & Cohen, L. Learning strategies as a mediator for motivation and a sense of achievement among students who study in MOOCs. *Education and Information Technologies, 22*(3), 1271–1290. https://doi.org/10.1007/s10639-016-9492-y

Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data: A model comparison perspective* (2nd ed.). Psychology Press.

Mayer, R. E., Fiorella, L., & Stull, A. (2020). Five ways to increase the effectiveness of instructional video. *Educational Technology Research and Development, 68*(3), 837-852. https://doi.org/10.1007/s11423-020-09749-6

Meyers, L. S., Gamst, G., & Guarino, A. J. (2016). *Applied multivariate research: Design and interpretation.* Sage.

Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive load approach. *Journal of Educational Psychology, 84*(4), 429–434. https://doi.org/10.1037/0022-0663.84.4.429

Pi, Z., Zhang, Y., Zhou, W., Xu, K., Chen, Y., Yang, J., & Zhao, Q. (2021). Learning by explaining to oneself and a peer enhances learners' theta and alpha oscillations while watching video lectures. *British Journal of Educational Technology, 52*(2), 659-679. https://doi.org/10.1111/bjet.13048

Poquet, O., Lim, L., Mirriahi, N., Dawson, S. (2018). Video and learning: A systematic review (2007–2017). In *Proceedings of the 8th international conference on learning analytics and knowledge* (pp. 151–160). ACM. https://doi.org/10.1145/3170358.3170376

Putnam, A. L., and Roediger III, H. L. (2013). Does response mode affect amount recalled or the magnitude of the testing effect? *Memory and Cognition, 41*(1), 36–48. https://doi.org/10.3758/s13421-012-0245-x

Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language, 60*(4), 437-447. https://doi.org/10.1016/j.jml.2009.01.004

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

Rittle-Johnson, B., & Loehr, A. M. (2017). Eliciting explanations: Constraints on when self-explanation aids learning. *Psychonomic Bulletin & Review, 24*(5), 1501-1510. https://doi.org/10.3758/s13423-016-1079-5

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

Roelle, J., Froese, L., Krebs, R., Obergassel, N., & Waldeyer, J. (2022). Sequence matters! Retrieval practice before generative learning is more effective than the reverse order. *Learning and Instruction, 80*, 1-12. https://doi.org/10.1016/j.learninstruc.2022.101634

Roelle, J., & Nückles, M. (2019). Generative learning versus retrieval practice in learning from text: The cohesion and elaboration of the text matters. *Journal of Educational Psychology, 111*(8), 1341-1361. https://doi.org/10.1037/edu0000345

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

Roy, M., & Chi, M. T. (2005). The self-explanation principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 271–286). Cambridge University Press. https://doi.org/10.1017/cbo9780511816819.018

Sadi, O., & Uyar, M. (2013). The relationship between self-efficacy, self-regulated learning strategies and achievement: A path model. *Journal of Baltic Science Education*, *12*(1), 21-33. https://doi.org/10.33225/jbse/13.12.21

Tabachnick, B. G., & Fidell, L. S. (2014). *Using multivariate statistics* (6th ed.). Pearson.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2,* 53-55. https://doi.org/10.5116/ijme.4dfb.8dfd

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: How writing and reading summaries of videos affects student learning, In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge* (pp. 216–225). https://doi.org/10.1145/3170358.3170382

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

Waldeyer, J., Heitmann, S., Moning, J., & Roelle, J. (2020). Can generative learning tasks be optimized by incorporation of retrieval practice? *Journal of Applied Research in Memory and Cognition, 9*(3), 355-369. https://doi.org/10.1016/j.jarmac.2020.05.001

Weinfurt, K. P. (1995). Multivariate analysis of variance. In Grimm, L. G. & P. R. Yarnold (Eds.), *Reading and understanding multivariate statistics* (pp. 245-276). American Psychological Association.

Wittrock, M. C. (2010). Learning as a generative process. *Educational Psychologist, 45*(1), 40–45. https://doi.org/10.1080/00461520903433554

Wylie, R., & Chi, M. T. H. (2014). The self-explanation principle in multimedia learning. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 413–432). Cambridge University Press. https://doi.org/10.1017/cbo9781139547369.021

Appendix A

Authors	Strategies	Materials and participants	Experimental design	Effects on learning outcomes
Fiorella & Kuhlmann (2020)	Drawing and teaching	Multimedia lesson on the human respiratory system, university students	2 × 2 between-subjects design: drawing (with vs. without) vs. teaching (with vs. without) In the condition with both strategies, strategies were used simultaneously.	• Retention: teaching + drawing > teaching (*d* = 0.96); teaching + drawing > drawing (*d* = 0.62); teaching + drawing > restudy (*d* = 1.15); drawing > restudy (*d* = 0.64) • Transfer: teaching + drawing > teaching (*d* = 0.59); teaching + drawing > restudy (*d* = 1.34); drawing > restudy (*d* = 0.89); teaching > restudy (*d* = 0.76)
Fiorella et al. (2021)	Sequencing: drawing and explaining	Multimedia lesson on the human respiratory system, university students	2 × 1 between-subjects design: draw-then-explain vs. explain-then-draw	• Explanation test: ns • Drawing test: ns • Transfer test: ns
Koh et al. (2018)	Retrieval practice (free recall) and teaching	Multimedia lesson on the Doppler effect, university students	4 x 1 between subjects design: control vs. retrieval practice vs. teaching (with teacher notes) vs. simultaneously using retrieval practice and teaching (without teacher notes)	• Comprehension: retrieval + teaching > control (*d* = 0.84); retrieval > control (*d* = 0.91); retrieval + teaching > teaching (*d* = 0.58); retrieval > teaching (*d* = 0.65)
Larsen et al. (2013)	Testing and self-explaining	A teaching session that covered four clinical topics and four weekly learning sessions, university students	4 x 1 between-subjects design: testing + self-explaining vs. testing vs. self-explaining vs. restudy In the condition with both strategies, strategies were used simultaneously.	• Retention and application: testing + self-explaining > self-explaining (*d* = 0.70); testing > self-explaining (*d* = 0.48); self-explaining > restudy (*d* = 0.68)
Roelle & Nückles (2019)	Retrieval practice (free recall) and explaining/giving examples	Experiment 1: expository text of high cohesion and elaboration, university students Experiment 2: expository text of low cohesion and elaboration, university students	2 x 2 between-subjects design: retrieval practice (with vs. without) vs. explaining/giving examples (with vs. without) In the explaining/giving examples condition, an open-book format was used.	Experiment 1: • Recognition: ns • Reproduction: ns • Comprehension: retrieval > restudy (*d* = 0.62) Experiment 2: • Reproduction: ns

Authors	Strategies	Materials and participants	Experimental design	Effects on learning outcomes
			In the condition with both strategies, strategies were used simultaneously.	• Comprehension: explaining/giving examples > explaining/giving examples with retrieval (*d* = 0.55) • Transfer: explaining/giving examples > explaining/giving examples with retrieval (*d* = 0.60) In both experiments, no effects of using multiple strategies were found.
Roelle et al. (2022)	Sequencing: retrieval practice (cued recall) and giving examples	Expository text, university students	3 × 2 between-subjects design: sequence (retrieval-before-giving examples vs. giving examples-before-retrieval condition vs. restudy) vs. timing of the posttest (immediate vs. delayed) For the giving examples condition, an open-book format was used.	• Learning gains cued recall: retrieval-before-giving examples > giving examples-before-retrieval (*d* = 0.56); retrieval-before-giving examples > restudy (*d* = 1.59) • Learning gains example generation: ns
Waldeyer et al. (2020)	Retrieval practice (free recall) and explaining/giving examples	Expository text, high school students	2 x 1 between-subjects design: explaining/giving examples (open-book format) vs. simultaneously using retrieval practice and explaining/giving examples (closed-book format) In the condition with both strategies, strategies were used simultaneously.	• Comprehension: ns No effects of using multiple strategies were found.