Do Video Modeling and Metacognitive Prompts Improve Self-Regulated Scientific Inquiry?

Yoana Omarchevska¹ · Andreas Lachner^{1,2} · Juliane Richter^{1,3} · Katharina Scheiter^{1,2}

Accepted: 14 December 2021 / Published online: 20 January 2022
© The Author(s) 2022

Abstract

Guided inquiry learning is an effective method for learning about scientific concepts. The present study investigated the effects of combining video modeling (VM) examples and metacognitive prompts on university students' ($N=127$) scientific reasoning and self-regulation during inquiry learning. We compared the effects of watching VM examples combined with prompts (VMP) to watching VM examples only, and to unguided inquiry (control) in a training and a transfer task. Dependent variables were scientific reasoning ability, hypothesis and argumentation quality, and scientific reasoning and self-regulation processes. Participants in the VMP and VM conditions had higher hypothesis and argumentation quality in the training task and higher hypothesis quality in the transfer task compared to the control group. There was no added benefit of the prompts. Screen captures and think aloud protocols during the two tasks served to obtain insights into students' scientific reasoning and self-regulation processes. Epistemic network analysis (ENA) and process mining were used to model the co-occurrence and sequences of these processes. The ENA identified stronger co-occurrences between scientific reasoning and self-regulation processes in the two VM conditions compared to the control condition. Process mining revealed that in the VM conditions these processes occurred in unique sequences and that self-regulation processes had many self-loops. Our findings show that video modeling examples are a promising instructional method for supporting inquiry learning on both the process and the learning outcomes level.

Keywords Scientific reasoning · Self-regulation · Video modeling · Metacognitive prompting · Epistemic network analysis · Process mining

✉ Yoana Omarchevska
y.omarchevska@iwm-tuebingen.de

¹ Leibniz-Institut Für Wissensmedien, Schleierstraße 6, 72076 Tübingen, Germany

² University of Tübingen, Tübingen, Germany

³ Macromedia University of Applied Sciences, Stuttgart, Germany

Improving scientific reasoning and argumentation is a central aim of science education (Engelmann et al., 2016; OECD, 2013). Consequently, science education has moved toward more inquiry-based learning approaches. Learning from inquiry can be more effective than direct instruction when appropriately guided (Lazonder & Harmsen, 2016). Typically, students use computer simulations to explore scientific concepts by testing hypotheses, conducting experiments, and evaluating data.

Inquiry learning can improve scientific reasoning by having students “act like scientists”, thereby improving their learning of the content and the corresponding scientific processes (Abd-El-Khalick et al., 2004). However, students might struggle with inquiry learning because they lack 1) scientific reasoning skills to conduct experiments or 2) self-regulation abilities, which are particularly important for navigating through complex learning environments like simulations. The present study tested the effectiveness of two types of guidance—video modeling examples and metacognitive prompts. The video modeling examples provided an integrated instruction of scientific reasoning and self-regulated learning. The metacognitive prompts aimed to further ensure the use of self-regulation processes by prompting students to monitor their scientific reasoning activities during inquiry. To our knowledge, this is the first study to develop an intervention aimed at simultaneously fostering scientific reasoning and self-regulation processes in an integrated way and test its effectiveness both at the process and product level (i.e., hypothesis and argumentation quality). To show the intervention’s effectiveness at the process level, we introduced two statistical methods to analyze the conjoint and sequential use of both types of processes that so far have been used only sparingly in educational research, namely, ENA and process mining.

Theoretical Framework

Scientific Reasoning and Argumentation

Scientific reasoning and argumentation skills are essential for comprehending and evaluating scientific findings (Engelmann et al., 2016; Pedaste et al., 2015). These skills refer to understanding how scientific knowledge is created, the scientific methods, and the validity of scientific findings (Fischer et al., 2014). Scientific reasoning and argumentation are defined as a set of eight epistemic activities, applicable across scientific domains (extending beyond the natural sciences, see Renkl, 2018 for a similar discussion)—problem identification, questioning, hypothesis generation, construction and redesign of artefacts, evidence generation, evidence evaluation, drawing conclusions, and communicating and scrutinizing (Fischer et al., 2014; Hetmanek et al., 2018). During problem identification a problem representation is built, followed by questioning, during which specific research questions are identified. Hypothesis generation is concerned with formulating potential answers to the research question, which are based on prior evidence and/or theoretical models. To test the generated hypothesis, an artefact can be constructed and later revised based on the evidence. Evidence is generated using controlled experiments, observations, or deductive reasoning to test the hypothesis. An important strategy for correct

evidence generation is the control-of-variables strategy (CVS; Chen & Klahr, 1999), which postulates that only the variable of interest should be manipulated, while all other variables are held constant. The generated evidence is evaluated with respect to the original theory. Next, multiple pieces of evidence are integrated to draw conclusions and revise the original claim.

Argumentation can be considered a consequence of scientific reasoning because the generated evidence is used to draw conclusions about scientific issues (Engelmann et al., 2016). We measured argumentation quality using the claim-evidence-reasoning (CER) framework which breaks down argumentation into a claim, evidence, and reasoning (McNeill et al., 2006). The claim answers the research question, the evidence is the data provided to support the claim, and the reasoning is the justification why the evidence supports the claim. Last, findings are scrutinized and communicated to a broader audience.

Students and adults often struggle with argumentation (Koslowski, 2012; Kuhn, 1991; McNeill, 2011) and scientific reasoning (de Jong & van Joolingen, 1998). However, scientific reasoning and argumentation can be improved with instruction and practice (Osborne et al., 2004), for example, using inquiry learning.

Computer-Supported Inquiry Learning

Students can use online simulations to actively learn about scientific concepts and the inquiry process (Zacharia et al., 2015). During inquiry, students apply some or all of the aforementioned scientific reasoning processes (van Joolingen & Zacharia, 2009). Using online simulations, students can conduct multiple experiments in a short amount of time and investigate concepts which are otherwise difficult to explore (e.g., evolution). More importantly, computer-supported inquiry learning environments provide unique opportunities for learning, like multiple representations and non-linear presentation of information (de Jong, 2006; Furtak et al., 2012).

Students' active engagement during inquiry learning can pose cognitive and metacognitive challenges for them (Azevedo, 2005; Scheiter & Gerjets, 2007). The (lack of) understanding of the scientific phenomenon or insufficient inquiry skills (e.g., inability to generate a testable hypothesis or designing an unconfounded experiment) can pose cognitive challenges. Moreover, students can experience metacognitive challenges because they need to self-regulate their inquiry process (Hadwin & Winne, 2001; Pintrich, 2000).

Self-Regulated Learning

The importance of metacognition (and self-regulation) for successful scientific reasoning was stressed more than 20 years ago (White & Frederiksen, 1998; Schunk and Zimmerman, 1998). Self-regulated learning is an active, temporal, and cyclical process (Zimmerman, 2013), during which learners set goals, monitor, regulate, and control their cognition, motivation, and behavior to meet their goals (Boekaerts, 1999; Pintrich, 1999). Metacognition is the cognitive component of self-regulated learning (Zimmerman & Moylan, 2009) and is predominantly concerned with

monitoring and regulation of learning (Nelson & Narens, 1990). Monitoring refers to students' ability to accurately judge their own learning. It provides the basis for regulation, that is, students' selection and application of learning strategies. Self-regulation is particularly important for successful inquiry learning (Chin & Brown, 2000; Kuhn et al., 2000; Omarchevska et al., 2021; Reid et al., 2003; White et al., 2009). For instance, students need to monitor whether they are manipulating the correct variables or how much data they need before drawing a conclusion. According to a fine-grained analysis of students' self-regulation and scientific reasoning processes monitoring during scientific reasoning activities was associated with higher argumentation quality (Omarchevska et al., 2021).

Because of the fundamental importance of accurate monitoring, we assessed metacognitive monitoring accuracy in relation to hypothesis and argumentation quality using retrospective confidence judgements (Busey et al., 2000). Moreover, we assessed students' academic self-concept and interest. Interest and academic self-concept are motivational factors that can influence self-regulation (Hidi & Ainley, 2008; Ommundsen et al., 2005). Interest is a psychological state with both affective and cognitive components that is also a predisposition to re-engage with the content in the future (Hidi & Renninger, 2006). Interest is positively associated with understanding, effort, perseverance (Hidi, 1990) and the maintenance of self-regulation (e.g., goal setting, use of learning strategies; Renninger & Hidi, 2019). Academic self-concept is a person's perceived ability in a domain (e.g., biology; Marsh & Martin, 2011) which is positively related to effort (Huang, 2011), interest (Trautwein & Möller, 2016), achievement (Marsh & Martin, 2011) and self-regulation strategies (Ommundsen et al., 2005). Therefore, we controlled for students' interest and academic self-concept.

Guidance During Computer-Supported Inquiry Learning

Guidance during inquiry can support students' learning both cognitively and metacognitively by tailoring the learning experience to their needs during specific phases of inquiry (Quintana et al., 2004). Guidance can be provided using process constraints, performance dashboards, prompts, heuristics, scaffolds, or direct presentation of information (de Jong & Lazonder, 2014). Furthermore, guidance should aim to also support self-regulated learning (Zacharia et al., 2015), as self-regulation has been shown to be important to successful inquiry learning (Omarchevska et al., 2021). Therefore, combining scientific reasoning and self-regulation instruction might be beneficial for teaching scientific reasoning. However, only few studies have invested whether supporting self-regulation during inquiry improves learning (Lai et al., 2018; Manlove et al., 2007, 2009). Last, more research on the effects of combining different types of guidance is needed (Lazonder & Harmsen, 2016; Zacharia et al., 2015). Therefore, we used video modeling examples to support scientific reasoning and self-regulation in an integrated way; in addition, metacognitive prompts were implemented to further support monitoring.

Video Modeling Examples

The rationale for using video modeling examples is rooted in theories of example-based learning during which learners acquire new skills by seeing examples of how to perform them correctly. Novice learners can benefit from studying a detailed step-by-step solution to a task before attempting to solve a problem themselves (Renkl, 2014; van Gog & Rummel, 2010). Studying worked examples reduces unnecessary cognitive load and frees up working memory resources so learners can build a problem-solving schema (Cooper & Sweller, 1987; Renkl, 2014). Example-based learning has been studied from a cognitive (cognitive load theory; Sweller et al., 2011) and from a social-cognitive perspective (social learning theory; Bandura, 1986). From a cognitive perspective, most research has focused on the effects of text-based worked examples, whereas social-cognitive studies have focused on (video) modeling examples (cf. Hoogerheide et al., 2014). Video modeling examples integrate features of worked examples and modeling examples (van Gog & Rummel, 2010) and they often include a screen recording of the model's problem-solving behavior combined with verbal explanations of the problem-solving steps (McLaren et al., 2008; van Gog, 2011; van Gog et al., 2009).

Video modeling examples can support inquiry and learning about scientific reasoning principles (Kant et al., 2017; Mulder et al., 2014). Watching video modeling examples before or instead of an inquiry task led to performing more controlled experiments, indicating that students can learn an abstract concept like controlling variables (CVS) using video modeling examples (Kant et al., 2017; Mulder et al., 2014). Outside the context of inquiry learning, using video modeling examples to train self-regulation skills (self-assessment and task selection) improved students' learning outcomes in a similar task (Kostons et al., 2012; Raaijmakers et al., 2018a, 2018b) but this outcome did not transfer to a different domain (Raaijmakers et al., 2018a).

Nevertheless, most studies have focused on either supporting scientific reasoning (e.g., CVS, Kant et al., 2017; Mulder et al., 2014) or self-regulation during inquiry learning (Manlove et al., 2007). Likewise, video modeling research has also focused on either supporting scientific reasoning (Kant et al., 2017; Mulder et al., 2014) or self-regulation (Raaijmakers et al., 2018a, 2018b). However, these studies have investigated scientific reasoning and self-regulation separately, whereas video modeling examples may be particularly suitable for integrating instruction of both scientific reasoning and self-regulated learning. Scientific reasoning principles can be easily demonstrated by showing how to conduct experiments correctly. Providing verbal explanations of the model's thought processes can be used to integrate self-regulated learning principles into instruction on scientific reasoning. For example, explaining the importance of planning for designing an experiment is one way to integrate these two constructs. Metacognitive monitoring can be demonstrated by having the model make a mistake, detect it, and then correct it (vicarious failure; Hartmann et al., 2020). In contrast to previous research focused on task selection and self-assessment skills (Kostons et al., 2012; Raaijmakers et al., 2018a, 2018b), we investigated the effectiveness of video modeling examples for training and transfer of other self-regulation skills—planning, monitoring, and control. Moreover, we

studied whether a video modeling intervention that integrates scientific reasoning and self-regulation instruction will improve inquiry learning. To ensure that participants engaged with the videos constructively (Chi & Wylie, 2014), we supplemented the video modeling examples with knowledge integration principles which involve “a dynamic process of linking, connecting, distinguishing, and structuring ideas about scientific phenomena” (Clark & Linn, 2009, p. 139). To further support self-regulated learning, we tested the effectiveness of combining video modeling examples with metacognitive prompts.

Metacognitive Prompting

Metacognitive prompts are instructional support tools that guide students to reflect on their learning and focus their attention on their thoughts and understanding (Lin, 2001). Prompting students to reflect on their learning can help activate their metacognitive knowledge and skills, which should enhance learning and transfer (Azevedo et al., 2016; Bannert et al., 2015). Metacognitive prompts support self-regulated learning by reminding students to execute specific metacognitive activities like planning, monitoring, evaluation, and goal specification (Bannert, 2009; Fyfe & Rittle-Johnson, 2016). Metacognitive prompts are effective for supporting students’ self-regulation (Azevedo & Hadwin, 2005; Dori et al., 2018) and hypothesis development (Kim & Pedersen, 2011) in computer-supported learning environments.

Even though providing support for self-regulated learning improves learning and academic performance on average (Belland et al., 2015; Zheng, 2016), some studies did not find beneficial effects of metacognitive support on learning outcomes (Mäeots et al., 2016; Reid et al., 2017). To understand why, it is necessary to consider the learning processes of students (Engelmann & Bannert, 2019). Process data could help determine whether students engaged in the processes as intended by the intervention or identify students who failed to do so. For instance, process data can provide further insights on the influence of prompts on the learning process (Engelmann & Bannert, 2019; Sonnenberg & Bannert, 2015).

Modeling Learning Processes

In the following, we will introduce two highly suitable methods for studying the interaction between scientific reasoning and self-regulation processes—epistemic network analysis (Shaffer, 2017) and process mining (van der Aalst, 2016). These methods go beyond the traditional coding-and-counting approaches by providing more insight into the co-occurrences and sequences of learning processes.

Epistemic Network Analysis

Epistemic network analysis (ENA; Shaffer, 2017) is a novel method for modeling the temporal associations between cognitive and metacognitive processes during learning. In ENA “the structure of connections among cognitive elements is more important than the mere presence or absence of these elements in isolation” (Shaffer

et al., 2016, p. 10). Therefore, it is essential to not only consider individual learning processes, but also preceding and following processes. ENA measures the structure and the strength of connections between processes, based on their temporal co-occurrence, and visualizes them in dynamic network models (Shaffer et al., 2016). The advantage of ENA is that the temporal patterns of individual connections can be easily captured and compared between individuals. In an exploratory think aloud study (Omarchevska et al., 2021), we found that students who were monitoring during scientific reasoning activities achieved higher argumentation quality than their peers who did not monitor using ENA. These findings demonstrated the added value of studying the temporal interaction between scientific reasoning and self-regulation processes and its effects on argumentation quality. The present study builds upon these findings by studying the effects of an intervention on learning processes as revealed not only by ENA, but also by process mining.

Process Mining

Process mining is a suitable method for modeling and understanding self-regulation processes (Bannert et al., 2014; Engelmann & Bannert, 2019; Roll & Winne, 2015). Process mining is a form of educational data mining, which uses event data to discover process models. Process models reveal the sequences between learning events, which provides insights into the sequential relationships between cognitive and metacognitive processes (Engelmann & Bannert, 2019).

In educational research, process mining was used to discover different student profiles and their learning processes in relation to their grades (Romero et al., 2010). Furthermore, process mining provided additional insights into the sequential structure of self-regulated learning processes (Bannert et al., 2014; Sonnenberg & Bannert, 2015, 2019). However, process mining techniques have not been used to model the relationship between scientific reasoning and self-regulated learning processes yet. Therefore, we used process mining to identify sequential relationships between scientific reasoning and self-regulation processes and combined it with ENA findings for a comprehensive analysis of the interaction between the two processes as each method has its unique benefits.

Process mining does not provide a statistical comparison between different process models, which is, however, offered within ENA. In contrast, ENA does not provide information about the direction of the relationship and does not consider when the same process is performed several times, whereas process mining provides information about the direction of the path and information about self-loops. To our knowledge, this is the first study to combine both methods and to use them to test the effects of educational interventions at the process level.

The Present Study

This study tested the effects of two types of guidance—video modeling examples and metacognitive prompts—on scientific reasoning performance and self-regulation during inquiry learning. Participants engaged in an inquiry training and transfer

task using two computer simulations. Screen captures and think aloud protocols were used to collect scientific reasoning and self-regulation process data, respectively. Effects of the intervention were expected to occur for 1) scientific reasoning ability as measured with a multiple-choice test, 2) scientific reasoning and self-regulation processes, and 3) the product of scientific reasoning, namely, the quality of the generated hypotheses and of the argumentation provided to justify decisions regarding the hypotheses. We preregistered (<https://aspredicted.org/vs43g.pdf>) the following research questions and hypotheses:

RQ1) Can video modeling and metacognitive prompts improve scientific reasoning ability?

In line with Kant et al. (2017), we hypothesized that students in the two VM conditions would have higher scientific reasoning posttest scores than the control group (H1a). Because of the benefits of providing metacognitive support (Azevedo & Hadwin, 2005), we hypothesized that the VMP condition would further outperform the VM condition (H1b).

RQ2) What are the immediate effects of video modeling and metacognitive prompts while working on an inquiry training task at the product level (hypothesis and argumentation quality) and process level (scientific reasoning and self-regulation)?

In line with Mulder et al. (2014), we hypothesized that students in the two VM conditions would have higher hypothesis and argumentation quality (H2a) than the control group in the training task. In line with Kim and Pedersen (2011), we hypothesized that the VMP condition would further outperform the VM condition (H2b).

RQ3) Do the effects of video modeling and metacognitive prompts on scientific reasoning products and processes transfer to a novel task?

In line with van Gog and Rummel (2010), we hypothesized that students in the two VM conditions would have higher hypothesis and argumentation quality (H3a) than the control group in the transfer task. In line with Bannert et al. (2015), we hypothesized that the VMP condition would outperform the VM condition (H3b).

Additionally, we explored the process models of participants' scientific reasoning and self-regulation processes in different conditions using ENA and process mining in the two tasks. Moreover, we explored participants' monitoring accuracy for hypothesis and argumentation quality in both tasks.

Method

Participants and Design

Participants were 127 university students from Southern Germany (26 males, $M_{age}=24.3$ years, $SD=4.81$). Participants had an academic background in science ($n=40$), humanities ($n=43$), law ($n=11$), social science ($n=20$), or other ($n=14$). Participation in the experiment was voluntary and informed consents were obtained from all participants. The study was approved by the local ethics committee (2019/031). The experiment lasted 1 h and 30 min and participants received a monetary reward of 12 Euros.

Fig. 1 The design and procedure of the experiment divided in three phases and the corresponding virtual experiment used in each phase

The experiment had a one-factorial design with three levels and participants were randomly placed in one of three conditions. In the first condition (VMP, $n=43$), participants watched video modeling examples (VM) before working with the virtual experiments and they received metacognitive prompts (P) during the training phase (see Fig. 1). In the second condition (VM, $n=43$), participants watched the same video modeling examples without receiving metacognitive prompts during the training phase. In the third condition (control, $n=41$), participants engaged in unguided inquiry task with the same virtual experiment that was used in the video modeling examples; however, they received neither video modeling instruction nor metacognitive prompts.

A priori power analysis using G*Power (Faul et al., 2007) determined the required sample size to be 128 participants (Cohen's $f=0.25$, power=0.80, $\alpha=0.05$) for contrast analyses. Effect size calculations were based on previous research using video modeling examples to enhance scientific reasoning (Kant et al., 2017). Data from one participant were not recorded due to technical issues, resulting in a sample size of 127.

Fig. 2 The Archimedes' Principle simulation. Copyright (2021) by Explore Learning. Reprinted with permission

Materials and Procedure

Phase 1—Instruction

We first assessed demographic information, conceptual knowledge, academic interest and self-concept (Fig. 1). During the instruction, participants either watched video modeling examples (intervention groups) or they engaged in an unguided inquiry learning task using the simulation *Archimedes' Principle* (control group, Fig. 2). In this simulation, a boat is floating in a tank of water. The boat's dimensions and weight and the liquid's density can be varied. When the boat sinks, the displaced liquid overflows in a cylinder. In this way, Archimedes' principle, which states that the upward buoyant force that is exerted on a body immersed in a fluid, is equal to the weight of the fluid that the body displaces, can be investigated.

In the two video modeling conditions, participants watched 3 non-interactive videos (each 3 min on average). The videos were screen captures recorded using Camtasia Studio which showed a female model's interactions with the simulation *Archimedes' principle*. The model was thinking aloud and explaining the different steps of scientific reasoning, but she was not visible in the videos.

To engage participants with the videos, knowledge integration principles were used (Clark & Linn, 2009). Before watching each video, participants' ideas about the topic of each video modeling example were elicited (e.g., "When conducting a scientific experiment, what is important to keep in mind before you start collecting

data?”). After watching each video, participants noted down the most important points and compared their first answer to what was explained in the video, which engaged them in reflection (Davis, 2000).

In the first video, the model explained *problem identification* and *hypothesis generation*. She explained how to formulate a research question and a testable hypothesis. Then, she developed her own hypothesis, which she later tested.

In the second video, the model explained *planning a scientific experiment* and the *control of variables strategy* (CVS). To demonstrate CVS, we used a coping model (van Gog & Rummel, 2010), who initially made a mistake by manipulating an irrelevant variable, which she then corrected, and explained that manipulating irrelevant variables can lead to confounded results, thereby also demonstrating *metacognitive monitoring*.

In the last video, *evidence generation*, *evidence evaluation*, and *drawing conclusions* were modeled by conducting an experiment to test the hypothesis. Data were systematically collected and presented in a graph. The model explained the importance of conducting multiple experiments to not draw conclusions prematurely, which also modeled *metacognitive monitoring* and *control*. Last, she evaluated the evidence and drew conclusions.

In the control condition, participants worked with the same virtual experiment used in the videos without receiving guidance. They answered the same research question as the model in the videos. To keep time on task similar between conditions, participants had 10 min to work on the task.

Phase 2—Training Task

Participants were first instructed to think aloud by asking them to say everything that comes to their mind without worrying about the formulation. Participants were given a short practice task (“Your sister’s husband’s son is your children’s cousin. How is he related to your brother? Please say anything you think out loud as you get to an answer.”). Participants watched a short video about photosynthesis, which served to re-activate their conceptual knowledge. Then, they solved the training inquiry task using the simulation *Photosynthesis* and an experimentation sheet. In *Photosynthesis* (see Fig. 3), the rates of photosynthesis (measured by oxygen production) are inferred by manipulating different variables (e.g., light intensity).

Participants were asked to answer the following research question: “How does light intensity influence oxygen production during photosynthesis?”. They wrote down their hypothesis, collected data using the simulation and answered the research question on the experimentation sheet. We asked participants to support their answers with evidence. Hypothesis and argumentation quality were coded from these answers. Participants made retrospective confidence judgments regarding their hypothesis (“How confident are you that you have a testable hypothesis?”) and their final answer (“How confident are you that you have answered the research question correctly?”).

Participants in the VM and in the control condition solved the task using only the *Photosynthesis* simulation and the worksheet. In the VMP condition, students additionally received 3 metacognitive prompts (see Table 1), which asked them to

Fig. 3 The Photosynthesis simulation. Copyright (2021) by Explore Learning. Reprinted with permission

Table 1 The metacognitive prompts used in the video modeling and prompting condition

Process	Prompt
Evidence Generation	How confident are you that you are manipulating the correct variable to test your hypothesis?
Evidence Generation	How confident are you that you have collected enough data to test your hypothesis?
Evidence Evaluation	How confident are you that you have interpreted your data correctly?

monitor specific scientific reasoning activities. Each prompt asked participants to rate their confidence on a scale from 0 to 100. The first two prompts were presented as pop-up messages during the training task after 3 and 9 min, respectively. The third prompt was visible after participants finished the training task and gave the option to go back and conduct more experiments.

Phase 3—Transfer Task

In the transfer task, all participants worked with the *Energy Conversion in a System* (see Fig. 4) simulation. First, participants read a short text about the law of conservation of energy which provided them with the necessary conceptual knowledge to use the simulation. In *Energy Conversion in a System*, participants could manipulate the quantity and the initial temperature of water in a beaker. Water is heated using a

Fig. 4 The Energy Conversion in a System simulation. Copyright (2021) by Explore Learning. Reprinted with permission

falling cylinder attached to a rotating propeller that stirs the water in a beaker. The mass and height of the cylinder could be adjusted. The change in the water's temperature is measured and energy is converted from one form to another.

The transfer task had an identical structure to the training task and was delivered through the experimentation sheet. Participants were asked to use the simulation to answer the question “How does changing the waters’ initial temperature and the water’s mass affect the change in temperature?”. Participants investigated the influence of two variables (water mass and water temperature) and noted down their results in a table with four columns (water mass, water initial temperature, water final temperature, change in temperature), which provided further guidance. Retrospective confidence judgments were provided for the hypothesis and the final answer. The task was the same in all conditions.

Measures

Conceptual Knowledge

Conceptual knowledge in photosynthesis (e.g., “What is the function of the chloroplasts?”) and energy conversion (e.g., “The law of conservation of energy states that...”) was assessed prior to the experiment using 5 multiple-choice items with 5 answer options for each topic. Each question had one correct answer and “I do not know” was one of the answer options. Both scales had low internal consistency

(photosynthesis, Cronbach's $\alpha=0.63$; energy conversion, Cronbach's $\alpha=0.15$), because they assessed prior understanding of independent facets related to photosynthesis and energy conversion. Therefore, computing internal consistency for such scales might not be appropriate (Stadler et al., 2021).

Academic Self-Concept and Interest in Science

The academic self-concept scale comprised 5 items rated on a Likert scale (Cronbach's $\alpha=0.93$) ranging from 1 (*I do not agree at all*) to 4 (*I completely agree*) (Grüß-Niehaus, 2010; Schanze, 2002). An example item of the scale is "I quickly learn new material in natural sciences.". Likewise, interest in science was assessed using a 5-item Likert scale (Cronbach's $\alpha=0.94$) ranging from 1 (*I do not agree at all*) to 4 (*I completely agree*) (Wilde et al., 2009). An example item of the scale is "I am interested in learning something new in natural sciences.".

Scientific Reasoning Ability

Scientific reasoning ability was assessed using 12 items from a comprehensive instrument (Hartmann et al., 2015; Krüger et al., 2020) that assessed the skills *research question formulation* (4 items), *hypothesis generation* (4 items), and *experimental design* (4 items). For each skill, we chose easy, medium, and difficult questions, based on data obtained by the authors of the instrument. Since the test was originally developed for pre-service science teachers, we chose items that did not rely on prior content knowledge. The questions matched the domains of the inquiry tasks (biology, physics). The scale had low internal consistency (Cronbach's $\alpha=0.31$), most likely because the test assessed three independent skills (cf. Stadler et al., 2021).

Scientific Reasoning Products

Hypothesis Quality To assess hypothesis quality, we developed a coding scheme which scored participants' hypotheses based on their testability (0–2) and correctness (0–2), adding up to a maximum score of 4, see Table 2. Due to the complexity of the coding scheme, we used consensus ratings (Bradley et al., 2007) for the scoring (initial inter-rater agreement: Krippendorff's $\alpha=0.65$). Two raters independently scored all hypotheses ($N=127$) and then discussed all disagreements until a consensus was reached.

Argumentation Quality Argumentation quality was assessed by coding participants' answers to the research questions. The quality of the claim, the evidence, and the reasoning were scored between 0 and 2, adding up to a maximum score of 6 (McNeill et al., 2006), see Table 3. We adapted the coding scheme from McNeill et al. (2006) to the context of our study. Participants were given an extra point when they evaluated their hypothesis in their final answer. Again, we used consensus ratings (initial agreement: Krippendorff's $\alpha=0.67$) for the scoring.

Table 2 Coding scheme used for assessing hypothesis quality

Level		
Component	0	1 2
Testability		
Criteria		
	The hypothesis is not testable, it is stated as a question, it does not make a prediction. The dependent or independent variables are not specified	The hypothesis is partially testable. The dependent and independent variables are stated but the direction of the relationship is not specified
Example from data	<i>Under what conditions does the plant produce its maximum amount of oxygen?</i>	<i>Light intensity influences oxygen production during photosynthesis</i>
Correctness		
Criteria	The hypothesis is incorrect or not stated at all	The hypothesis is partially correct (one variable is correct, the other is incorrect or irrelevant)
Example from data	<i>Plants need light for photosynthesis</i>	<i>Temperature has an effect on oxygen production. (irrelevant variable)</i>

Note. The examples given in the table are taken from the training task. The same coding scheme was applied for the transfer task

Table 3 Coding scheme used for assessing argumentation quality

Level		1		2
Component	0			
Claim				
Criteria				
Example from data	Does not make a claim or makes an inaccurate claim <i>The example shows that a lot of oxygen was produced at full light intensity (100%). (inaccurate claim)</i>	Makes an accurate but incomplete claim. Includes irrelevant variables <i>Depending on the temperature, the curve of oxygen production runs longer or shorter linearly before it stagnates. (irrelevant variable: temperature)</i>	Makes an accurate and complete claim <i>If the light intensity is increased above 40%, the oxygen production remains unchanged at 30 mL/h</i>	
Evidence				
Criteria	Does not provide evidence or only provides inappropriate evidence	Appropriate and insufficient or inappropriate evidence	Provides appropriate evidence and no inappropriate evidence	
Example from data	As light intensity increases, oxygen production increases up to a certain point. (vague evidence)	<i>oxygen production increases when certain temperature (perfect 24–28), right light intensity (50–60%), quite high CO₂ level (from 420 ppm) (appropriate and inappropriate evidence)</i>	<i>The light intensity influences the oxygen flow linearly up to a value of 37% under constant conditions</i>	
Reasoning				
Criteria	Does not provide reasoning or only provides reasoning that does not link evidence to the claim	Repeats evidence that links it to the claim, uses some scientific principles but insufficient	Accurate and complete reasoning that links evidence to claim, mentions scientific principles	
Example from data	<i>The initial temperature has an influence on the temperature at the end of the water and on the mass</i>	<i>With a light intensity of 60% we have a maximum O₂ production</i>	<i>The greater the mass of the water that is to be heated, the lower is the temperature increase, since the height energy, which is converted into heat, remains the same, but is distributed over more water molecules. This happens independently of the initial temperature</i>	

Scientific Reasoning and Self-Regulation Processes

We used screen captures and think aloud protocols to assess scientific reasoning and self-regulation processes, which were coded using Mangold INTERACT® (Mangold International GmbH, Arnstorf, Germany; version 9.0.7). Using INTERACT, the audio (think aloud) and video (screen captures) can be coded simultaneously. First, two experienced raters independently coded 20% of the videos ($n=25$) and reached perfect agreement ($\kappa=1.00$); therefore, each rater coded half of the remaining videos. Due to technical issues with the audio recording, the process data analysis was conducted on a smaller sample ($N=88$, $n_{\text{VMP}}=29$, $n_{\text{VM}}=37$, $n_{\text{control}}=22$).

The raters used the coding scheme in Table 4, which was previously used by Omarchevska et al. (2021) to code scientific reasoning, self-regulation, and the use of cognitive strategies. Scientific reasoning processes were coded from both data sources, whereas self-regulation processes and use of cognitive strategies were coded from the think aloud protocols only. Regarding scientific reasoning, we focused on the epistemic activities *problem identification*, *hypothesis generation*, *evidence generation*, *evidence evaluation*, *drawing conclusions* (Fischer et al., 2014). As measures of self-regulation, we coded the processes of *planning* (goal setting), *monitoring* (goal progress, information relevance), and *control*. We coded the use of cognitive strategies, namely, *activation of prior knowledge* and *self-explaining*.

Monitoring Accuracy

Monitoring accuracy was measured by calculating bias scores, based on the match between the confidence judgements scores and the corresponding performance scores for hypothesis and argumentation quality (4 monitoring accuracy scores per participant). Since confidence judgements scores ranged from 0 to 100 and hypothesis and argumentation quality scores ranged from 0 to 6, they were rescaled to range between 0 and 100. Monitoring accuracy was computed by subtracting the hypothesis and argumentation quality scores from the confidence judgment scores (Baars et al., 2014). Positive scores indicate overestimation, negative scores indicate underestimation, and scores close to zero indicate accurate monitoring (Baars et al., 2014).

Data Analyses

To test our preregistered hypotheses, we used contrast analyses: Contrast 1 (0.5, 0.5, -1) compared the VMP and the VM conditions to the control condition and Contrast 2 (1, -1, 0) compared the VMP and the VM conditions to each other. We applied the Bonferroni correction for multiple tests which resulted in $\alpha=0.025$ for all contrast analyses. Benchmarks for effect sizes were: $\eta^2=0.01$, 0.06, 0.14 and $d=0.20$, 0.50, 0.80 for small, medium, and large effects, respectively.

We first compared the groups in the scientific reasoning ability posttest using contrast analysis (H1a, H1b) with scientific reasoning ability pretest as a covariate. We then applied contrast analyses to compare hypothesis and argumentation quality in the training (H2a, H2b) and the transfer task (H3a, H3b). We explored monitoring

accuracy in the two tasks using one-way ANOVAs. To explore the training and transfer effects on the process level, we used ENA and process mining.

Epistemic Network Analysis

ENA is a modeling tool which quantifies the strength of co-occurrence between codes within a conversation (Shaffer, 2017). A conversation is defined as a set of lines which are related to each other. ENA quantifies the co-occurrences between different codes in a conversation and visualizes them in network graphs. Hence, the strength of the co-occurrences can be visually and statistically compared between groups. Since the number of codes may vary, the networks are first normalized before they are subjected to a dimensional reduction. ENA uses singular-value decomposition to perform dimensional reduction, which produces orthogonal dimensions that maximize the variance explained by each dimension (Shaffer et al., 2016). The position of the networks' centroids, which correspond to the mean position of all points in the network, can be compared. The strength of individual connections can be compared using network difference graphs, which subtract the corresponding connection weights in different networks.

We used the ENA Web Tool (version 1.7.0) (Marquart et al., 2018) to compare the epistemic networks of participants in the VMP and VM conditions to the control condition in the training and the transfer tasks. Participants served as units of analysis and the two tasks as conversations. We used the coded process data from the screen captures and think aloud protocols as input for the analysis. We used the codes problem identification, hypothesis generation, evidence generation, evidence evaluation, drawing conclusions, planning, monitoring, control, and self-explaining (see Tables 6 and 7 for the training and transfer tasks, respectively). Due to the very low frequency of activation of prior knowledge, this code was excluded from all process analyses. Due to the low frequency of planning in the control condition, planning was also excluded from the analysis of the control condition in both tasks.

Process Mining

We used process mining to model the sequences of scientific reasoning and self-regulation processes in different conditions. We used the HeuristicsMiner algorithm (Weijters et al., 2006), to mine the sequence in which participants engaged in these processes during the training and the transfer tasks, which was implemented in the ProM framework version 5.2 (Verbeek et al., 2010). The HeuristicsMiner algorithm is well suited for educational data mining, because it deals well with noise and presents the most frequent behavior found in an event log without focusing on specifics and exceptions (i.e., low frequent behavior) (Weijters et al., 2006). The dependencies among the processes in an event log are represented in a heuristic net.

The heuristic net indicates the dependency and frequency of a relationship between two events (Weijters et al., 2006). In the heuristic net, the boxes represent the processes and the arcs connecting them represent the dependency between them. Dependency (0–1) represents the certainty between two events, with values closer to 1 indicating stronger dependency. The frequency reflects how often a transition

Table 4 Coding scheme used for assessing scientific reasoning, self-regulation, and cognitive strategies processes

Major category	Source	Definition or example from data
Scientific Reasoning		
Problem Identification	TA	"I need to investigate how temperature influences O_2 production."
Hypothesis Generation	TA	"If I increase light intensity, then oxygen production will increase."
Evidence Generation	Video	Participant manipulates the correct, incorrect, or irrelevant variable, use of CVS
Evidence Evaluation	Video + TA	"From 45% onwards the curve isn't rising anymore."
Drawing Conclusions	Video + TA	"This is only true up until a light intensity of 40%."
Self-Regulated Learning		
Planning	TA	"First, I will increase light intensity."
Goal setting	TA	"I will set the light intensity to 0% and increase it gradually by 10%."
Monitoring	TA	"Now I understand it."
Goal progress	TA	"Okay... I do not know yet about the CO_2 level."
Information Relevance	TA	"The temperature and the CO_2 level should remain constant."
Control	TA	"Now I will try and keep the temperature constant while increasing the mass."
Cognitive Strategies		
Activation of Prior Knowledge	TA	"I have to remember the video that was presented to me earlier."
Self-explanation	TA	"Which means that the initial temperature does not have any impact."

Note. TA think aloud, Video Screen capture. The examples are from the training and the transfer tasks

between two events occurred. The dependency is shown on the top, whereas the frequency is shown below it on each arrow. An arc is pointing back to the same box indicates a self-loop, showing that a process was observed multiple times in a row (Sonnenberg & Bannert, 2015). To ease generalizability to other studies using the HeuristicMiner (e.g., Engelmann & Bannert, 2019; Sonnenberg & Bannert, 2015), we kept the recommended default threshold values, namely, dependency threshold=0.90, relative-to-best-threshold=0.05, positive observations threshold=10. The dependency threshold determines the threshold for including dependency relations in the output model, the positive observations threshold determines the minimum observed sequences required to be included in the output model (Sonnenberg & Bannert, 2015; Weijters et al., 2006). For a detailed description of the HeuristicsMiner, the reader is referred to Weijters et al. (2006) and Sonnenberg and Bannert (2015). We used the same process data and exclusion criteria as in the ENA.

Results

Control Variables

A MANOVA with the control variables (conceptual knowledge, interest, academic self-concept) revealed no differences between conditions, $F < 1$, see Table 5 for descriptive statistics.

Scientific Reasoning Ability

A contrast analysis with scientific reasoning ability (pretest) as a covariate revealed no group differences in the posttest scores (Contrast 1, $\beta = 0.04$, $p = 0.57$, $d = 0.22$; Contrast 2, $\beta = -0.02$, $p = 0.76$, $d = 0$). Thus, there was no support for H1a and H1b.¹

Effects on the Training Task

Product Level Contrast 1 showed that the VMP and VM conditions had higher hypothesis quality, $t(124) = 2.60$, $p = 0.010$, $d = 0.49$, and higher argumentation quality, $t(124) = 2.75$, $p = 0.007$, $d = 0.52$, than the control group, see Fig. 5. Contrast 2 showed no significant differences between the VMP and VM in hypothesis quality, $p = 0.75$, $d = 0.02$, and argumentation quality, $p = 0.91$, $d = 0.02$. These findings indicate that while video modeling improved scientific inquiry at the product level (H2a), there was no significant added benefit of the metacognitive prompts in the training task (H2b). The groups did not differ in monitoring accuracy regarding hypothesis quality, $p = 0.20$, $\eta^2 = 0.03$, and argumentation quality, $p = 0.20$, $\eta^2 = 0.03$.

¹ Additional analyses revealed that time on task was significantly shorter for the posttest, $F(1, 121) = 475.21$, $p < .001$, Wilk's $\Lambda = .203$, $\eta_p^2 = .79$, suggesting that participants clicked through the posttest very quickly. Therefore, this finding should be interpreted with caution.

Table 5 Descriptive statistics (and SDs) for the control and dependent variables, and monitoring accuracy (bias)

Variable	VMP (<i>n</i> =43)	VM (<i>n</i> =43)	Control (<i>n</i> =41)
Conceptual knowledge training task (0–5)	3.90 (1.41)	3.84 (1.27)	4.10 (1.09)
Conceptual knowledge transfer task (0–5)	1.07 (0.80)	1.19 (0.79)	1.24 (0.70)
Academic self-concept in science (1–4)	1.34 (0.76)	1.20 (0.78)	1.32 (0.81)
Interest in science (1–4)	1.92 (0.87)	1.68 (0.80)	1.85 (0.82)
Scientific reasoning ability pretest (0–12)	5.41 (2.03)	5.30 (1.75)	5.17 (1.94)
Scientific reasoning ability post-test (0–12)	5.32 (2.19)	5.33 (1.76)	5.02 (2.19)
Hypothesis quality training task (0–4)	3.28 (1.33)	3.30 (1.05)	2.29 (1.54)
Argumentation quality training task (0–6)	3.72 (2.05)	3.70 (1.93)	2.51 (1.78)
Hypothesis quality transfer task (0–4)	1.98 (1.21)	2.10 (1.26)	1.33 (1.20)
Argumentation quality transfer task (0–6)	2.57 (1.89)	2.81 (1.68)	3.02 (1.21)
Bias hypothesis training task (0–100)	-6.16 (33.34)	-3.95 (27.97)	8.71 (43.13)
Bias argumentation training task (0–100)	11.01 (34.79)*	12.79 (34.72)*	23.83 (35.80)*
Bias hypothesis transfer task (0–100)	4.30 (33.98)	8.27 (38.74)	29.01 (31.66)*
Bias argumentation transfer task (0–100)	15.78 (35.49)*	22.50 (32.77)*	27.58 (23.47)*

* indicates a significant deviation from zero as revealed in a one-sample *t*-test ($p < .05$), Cohen's *d* ranges from 0.20 to 1.18 for significant differences

Process Level We used ENA and process mining to model the connections between scientific reasoning and self-regulation processes. We first compared the VMP and VM conditions to the control condition and then the two VM conditions to each other. The frequencies of the processes are provided in Table 6.

ENA: VMP and VM vs. Control. Along the X axis, a two sample *t*-test showed that the position of the control group centroid ($M=0.33$, $SD=0.76$, $N=22$) was significantly different from the VMP and VM centroid ($M=-0.11$, $SD=0.84$, $N=66$; $t(39.23)=-2.32$, $p=0.03$, $d=0.54$). Along the Y axis, the position of the control group's centroid ($M=-0.21$, $SD=0.74$, $N=22$) was not significantly different from the VMP and VM centroid ($M=0.07$, $SD=0.73$, $N=66$; $t(35.90)=-1.52$, $p=0.14$, $d=0.37$).

Thicker green lines in Fig. 6 illustrate stronger co-occurrences of Monitoring, Planning and Evidence Evaluation, Evidence Generation, Hypothesis Generation, and Drawing Conclusions in the two VM conditions than in the control condition. Participants in the control condition (purple) were Self-Explaining during Evidence Generation and Evidence Evaluation more often than in the VM conditions. The difference between the centroids' positions on the X-axis results from stronger connections between scientific reasoning and self-regulation processes in the VMP and VM networks. Taken together, participants in the VM conditions were self-regulating during scientific reasoning activities more frequently than participants in the control condition.

ENA: VMP vs. VM. Two sample-*t* tests revealed no differences between the position of the VMP centroid compared to the VM control along either the X axis,

Fig. 5 The immediate training effects of video modeling (VM) and metacognitive prompts (P) on hypothesis and argumentation quality. Error bars show standard errors

$t(55.53)=0.52$, $p=0.60$, $d=0.13$) or the Y axis, $t(59.54)=-0.42$, $p=0.67$, $d=0.11$. Thus, the epistemic networks of the groups with video modeling were similar (for details see online supplementary materials).

Process Mining. The process model of the two video modeling conditions in the training task (Fig. 7) illustrates a very strong dependency between Problem Identification and Hypothesis Generation. Participants started their inquiry by identifying the problem and then generating a hypothesis to investigate it. Next, there were strong sequential relationships between Evidence Generation and Planning. More frequently, participants were planning before evidence generation. A strong reciprocal relationship indicates that participants were also planning after evidence generation. Similar reciprocal relationships are observed between Evidence Generation, Self-Explaining and Evidence Evaluation, and between Drawing Conclusions and Evidence Generation. Monitoring and Control were not related to specific scientific reasoning processes and both processes had strong self-loops, which indicates that these processes were performed sequentially several times. This finding implies that Monitoring and Control did not have high enough dependencies to other events to be included in the process model. This could also indicate that Monitoring and Control were not related to one other specific process, but rather they were (weakly) connected to several other (scientific reasoning) processes.

Table 6 The total frequencies (and percentages) of process events per condition in the training task

	VMP (<i>n</i> = 29)	VM (<i>n</i> = 37)	Control (<i>n</i> = 22)
Problem Identification	20 (0.97%)	28 (0.95%)	17 (0.65%)
Hypothesis Generation	19 (0.92%)	32 (1.09%)	9 (0.35%)
Evidence Generation	1797 (86.69%)	2495 (84.95%)	2339 (89.9%)
Evidence Evaluation	133 (6.42%)	196 (6.67%)	124 (4.77%)
Drawing Conclusions	42 (2.03%)	66 (2.24%)	31 (1.19%)
Planning	13 (0.63%)	12 (0.41%)	2 (0.07%)
Monitoring	30 (1.45%)	54 (1.84%)	28 (1.07%)
Control	9 (0.43%)	31 (1.06%)	11 (0.42%)
Self-explaining	9 (0.43%)	18 (0.61%)	41 (1.58%)
Activation of Prior Knowledge	1 (0.05%)	5 (0.17%)	0
Total	2073	2937	2602

In contrast, in the control condition (Fig. 8), Problem Identification, Hypothesis Generation, and Monitoring were disconnected from other processes and no self-loops were observed. Similar to the two video modeling conditions, reciprocal links between Evidence Generation and Drawing Conclusions and between Self-Explaining and Evidence Evaluation were discovered. However, in contrast with the two video modeling conditions, the reciprocal links between Self-Explaining and Evidence Evaluation were disconnected from Evidence Generation.

Effects on the Transfer Task

Product Level Contrast 1 indicated that VMP and VM had higher hypothesis quality than the control condition, $t(124)=3.06$, $p=0.003$, $d=0.58$, but they did not differ significantly from each other (Contrast 2, $t(124)=-0.22$, $p=0.83$, $d=0.05$), see Fig. 9. The three groups did not differ in argumentation quality (Contrast 1, $t(124)=0.09$, $p=0.93$, $d=0.02$; Contrast 2, $t(124)=1.09$, $p=0.28$, $d=0.22$). These findings provide partial support for H3a in that video modeling helped students to generate high-quality hypotheses during scientific inquiry but did not improve their argumentation. There was no significant added benefit of metacognitive prompts in the transfer task (H3b).

The three groups differed significantly in monitoring accuracy regarding hypothesis quality, $F(2, 122)=5.68$, $p=0.004$, $\eta^2=0.09$. Post hoc comparisons with Bonferroni corrections indicated that participants in the control condition overestimated their hypothesis quality, compared to the VMP, $p=0.006$, and the VM condition, $p=0.028$. There were no differences in monitoring accuracy for argumentation quality, $p=0.20$, $\eta^2=0.02$.

Process Level Descriptive statistics are reported in Table 7.

ENA: VMP and VM vs. Control. Two sample-*t* tests revealed no differences between the position of the VMP and VM centroid compared to that of the

Fig. 6 The epistemic network difference between the VMP and VM (left, green) conditions and the control (right, purple) condition in the training task

Fig. 7 The process model of the VMP and VM conditions combined ($n = 66$) in the training task. The top number (referring to the arrows) represents the dependency (0–1) and the number below represents the frequency of each sequence. The numbers in the boxes represent the frequency of each process

Fig. 8 The process model of the control condition ($n=22$) in the training task. The top number (referring to the arrows) represents the dependency (0–1) and the number below represents the frequency of each sequence. The numbers in the boxes represent the occurrence of each process

control along either the X axis, $t(34.29)=-0.63$, $p=0.53$, $d=0.16$, or the Y axis, $t(34.90)=-1.43$, $p=0.16$, $d=0.36$. Thus, the epistemic networks of the groups with and without video modeling were similar (for details see online supplementary materials).

ENA: VMP vs. VM. Two sample t -tests showed no significant differences between the centroids of the VMP and VM conditions along the X, $t(57.37)=-1.62$, $p=0.11$, $d=0.40$, and the Y axis, $t(54.76)=-0.43$, $p=0.67$, $d=0.11$. These findings indicate that there were no significant differences between the epistemic networks of participants in the two video modeling conditions in the transfer task (for details see online supplementary materials).

Discussion

The present study investigated the effectiveness of video modeling examples and metacognitive prompts for improving scientific reasoning during inquiry learning with respect to students' scientific reasoning ability, self-regulation and scientific reasoning processes, and hypothesis and argumentation quality. We used two types of process analyses, ENA and process mining, to illustrate the interaction between self-regulated learning and scientific reasoning processes.

Our findings on the product level provide partial support for our hypotheses that video modeling improved hypothesis and argumentation quality in the training task (H2a) and hypothesis quality in the transfer task (H3a). Likewise, on the process level, we found more sequential relationships between scientific reasoning and self-regulated learning processes in the two video modeling conditions than in the control condition. We found no added benefit of the metacognitive prompts on either the product (H2b, H3b) or the process level.

We observed no effects on scientific reasoning ability (H1a, H1b), which is likely because of the identical pre- and posttest items. Participants' behavior during the posttest and the significantly shorter time-on-task suggested that

Fig. 9 The transfer effects of video modeling (VM) and metacognitive prompts (P) on hypothesis and argumentation quality. Error bars show standard errors

many participants rushed through the posttest and did not attempt to solve it again, thereby attesting to the importance of identifying alternative, more process-oriented measures for scientific reasoning. Therefore, we refrain from drawing conclusions about the influence of video modeling examples and metacognitive prompts on scientific reasoning ability and earmark this aspect of our study as a limitation.

Theoretical Contributions

First, video modeling examples were beneficial for improving participants' hypothesis and argumentation quality in the training task and for improving hypothesis quality in the transfer task. Therefore, video modeling examples can support students during inquiry learning and positively affect their hypothesis quality prior to the inquiry tasks and their argumentation quality. Previous research found positive effects of video modeling examples on performance and inquiry processes, compared to solving an inquiry task (Kant et al., 2017) and our findings extend them to hypothesis and argumentation quality. This provides further support of the benefits of observational learning (Bandura, 1986) for teaching complex procedural skills like scientific reasoning.

Second, next to learning outcome measures, we provide a fine-grained analysis of students' scientific reasoning and self-regulation processes. Prior research has stressed the importance of self-regulation during complex problem-solving activities (e.g., Azevedo et al., 2010; Bannert et al., 2015) like scientific reasoning (e.g.,

Table 7 The total frequencies (and percentages) of process events per condition in the transfer task

	VMP (<i>n</i> = 29)	VM (<i>n</i> = 37)	Control (<i>n</i> = 22)
Problem Identification	15 (1.06%)	29 (1.59%)	13 (1.15%)
Hypothesis Generation	11 (0.78%)	35 (1.92%)	9 (0.79%)
Evidence Generation	1037 (73.4%)	1208 (66.12%)	806 (71.39%)
Evidence Evaluation	235 (16.64%)	358 (19.59%)	192 (17%)
Drawing Conclusions	43 (3.05%)	67 (3.67%)	34 (3.01%)
Planning	8 (0.6%)	9 (0.5%)	2 (0.18%)
Monitoring	26 (1.84%)	52 (2.85%)	28 (2.48%)
Control	17 (1.2%)	35 (1.92%)	11 (0.87%)
Self-explaining	18 (1.28%)	33 (1.81%)	34 (3.01%)
Activation of Prior Knowledge	2 (0.14%)	1 (0.05%)	0
Total	1412	1827	1129

Manlove et al., 2009; Omarchevska et al., 2021; White et al., 2009). Our findings illustrate the importance of self-regulation processes during scientific reasoning activities and the role of video modeling for supporting this relationship. Specifically, students who watched video modeling examples were monitoring and planning during evidence generation and evidence evaluation more frequently than the control group. We contribute to the literature on self-regulated inquiry learning by showing that an integrated instruction of self-regulation and scientific reasoning resulted in more connections between scientific reasoning processes, more self-regulation, and higher hypothesis and argumentation quality. However, to ensure that an integrated instruction of scientific reasoning and self-regulation is more beneficial than an isolated one, future research should compare video modeling examples that teach self-regulation and scientific reasoning separately to our integrated instruction. Furthermore, we primarily relied on (meta)cognitive aspects of self-regulated learning. Future research should also investigate motivational influences (Smit et al., 2017) which were shown to be relevant for learning from video modeling examples (Wijnia & Baars, 2021).

Modeling the sequential use of scientific reasoning processes highlights the value of a process-oriented perspective for inquiry learning. ENA showed a densely connected network of scientific reasoning and self-regulation processes for students in the experimental conditions. This corroborates previous findings on the relationship between scientific reasoning and self-regulation processes and argumentation quality (Omarchevska et al., 2021) and extends it to the influence on hypothesis quality as well. The video modeling examples in this study were designed based on our previous findings regarding self-regulating during scientific reasoning, which further highlights the value of video modeling examples as a means of providing such an integrated instruction. Video modeling examples were successful in teaching self-regulation (self-assessment and task selection skills; Raaijmakers et al., 2018a, 2018b) or scientific reasoning (CVS; Kant et al., 2017; Mulder et al., 2014). Our findings extend these benefits for planning, monitoring, and the epistemic activities

regarding scientific reasoning (Fischer et al., 2014). Raaijmakers et al., (2018a, 2018b) showed mixed findings regarding transfer. In our case, our intervention transferred only regarding hypothesis quality, which was explicitly modeled in the videos, but did not enhance argumentation quality. The combination of process and learning outcome analyses demonstrated that 1) modeling hypothesis generation principles resulted in higher hypothesis quality and 2) integrating self-regulation and scientific reasoning instruction resulted in a conjoint use of scientific reasoning and self-regulation processes for students who watched the video modeling examples.

Our findings confirm the value of inquiry models (e.g., Fischer et al., 2014; Klahr & Dunbar, 1988; Pedaste et al., 2015) to not only describe the inquiry processes but to also inform interventions, as it was done in this study. At the same time, our findings suggest that current theoretical models describing scientific inquiry as a set of cognitive processes (Fischer et al., 2014; Klahr & Dunbar, 1988; Pedaste et al., 2015) need to be augmented with respect to metacognition. A framework integrating metacognition in the context of online inquiry was proposed by Quintana et al. (2005). The framework is focused on supporting metacognition in the context of information search and synthesis to answer a research question. The present study provides further evidence with respect to planning, monitoring, and control as important self-regulation processes for scientific reasoning and inquiry learning. Moreover, we tested the cyclical assumption of self-regulation, meaning that performance on one task can provide feedback for the learning strategies used in a follow-up task (Panadero & Alonso-Tapia, 2014). Our process analyses showed that increasing one component of self-regulated learning (e.g., monitoring) in the instruction can result in recursive use of that component in similar learning situations. This results in near transfer from applying the concepts taught in the video modeling examples to a new context (training task), but also in medium transfer, since the transfer task involved a new context and an additional challenge of manipulating a second variable. The cyclical nature of self-regulated learning is often assumed (Zimmerman, 2013), but it is rarely tested in successive learning experiences.

Methodological Contributions

We used two innovative process analyses—ENA and process mining—to investigate how scientific reasoning and self-regulation processes interacted in students who either watched video modeling examples or engaged in unguided inquiry learning prior to the inquiry tasks. Previous research on scientific reasoning using process data often used coding-and-counting methods, which use process frequencies to explain learning outcomes (e.g., Kant et al., 2017). A drawback of coding-and-counting methods is that when the frequencies of processes are analyzed in isolation, important information about the relationships between processes is lost (Csanadi et al., 2018; Reimann, 2009). For example, Kant et al. (2017) counted and compared the number of controlled and confounded experiments between conditions. While this analysis yielded important information about inquiry learning, it did not provide information about how participants engaged in these processes. Learning processes do not occur in isolation and studying their co-occurrence and sequential

flow can show how they are related to each other and provide a more comprehensive view of learning (Csanadi et al., 2018; Shaffer, 2017). Therefore, we use ENA and process mining to model the sequential relationships between scientific reasoning and self-regulation processes. Second, we integrated findings from both methods which can help to overcome the drawbacks of each method. Such approach is of interest to researchers analyzing learning processes as mediators between learning opportunities offered and students' learning outcomes. We illustrate the potential of advanced methods that go beyond the isolated frequencies of single processes without accounting for their interplay with other processes.

ENA statistically compares the position of the centroids and the strength of co-occurrences between conditions, whereas such comparison is difficult using process mining (Bolt et al., 2017). ENA found no differences between the two video modeling conditions; therefore, we compared the strength of connections among different processes between the video modeling conditions to the control condition. The two video modeling conditions significantly differed from the control condition in the training task and these differences were displayed in the epistemic networks of the two groups—scientific reasoning processes co-occurred more frequently with self-regulation processes in the video modeling conditions than the control condition.

ENA revealed that planning co-occurred with evidence generation more frequently in the experimental conditions, whereas process mining identified that more frequently, planning occurred prior to evidence generation, which could not be determined using ENA. Planning before generating evidence during inquiry learning is more beneficial because it helps students to first consider what experiment they want to conduct, which variables they want to manipulate, and to avoid randomly manipulating variables to test their hypothesis. The importance of planning during evidence generation was explained in the video modeling examples. Showing that participants applied this concept during their inquiry could explain the effectiveness of the video modeling examples. Furthermore, ENA showed stronger relationships between drawing conclusions and evidence generation in the experimental conditions. Process mining found the same reciprocal relationship and showed that, more frequently, evidence generation was followed by drawing conclusions. This relationship corresponds to scientific reasoning (e.g. Fischer et al., 2014) and inquiry (Pedaste et al., 2015) frameworks. In conclusion, ENA provides information about the strength of the co-occurrence between different processes, whereas process mining provides information about the direction.

Process mining identifies self-loops, indicating that the same process is performed several times in a row, whereas ENA does not consider loops. The process models identified monitoring self-loops but no strong dependencies between monitoring and other processes. In ENA, monitoring was connected to several scientific reasoning activities in the video modeling conditions. In the control condition, monitoring had no self-loops and it was disconnected from other activities; likewise, no relationships between monitoring and other processes were observed in the ENA. In conclusion, our findings show that a combination of both methods provides a more comprehensive analysis of the interaction between scientific reasoning and self-regulation processes. An integration of the findings of both approaches can complement the drawbacks of each method and provide information about global

differences between the groups (ENA), the strength of co-occurrence between specific processes (ENA), the specific sequence of relationships (process mining), and identify self-loops of individual processes (process mining).

Educational Implications

Based on our findings we can recommend the use of video modeling examples to teach scientific reasoning. Students who watched video modeling examples generated higher-quality hypotheses in a training and a transfer task, and higher-quality arguments in the training task, compared with students who engaged in unguided inquiry learning. Furthermore, video modeling examples enhanced self-regulation during scientific reasoning activities. Students who watched video modeling examples were planning and monitoring more frequently their scientific reasoning processes than students in the control group. The benefits were found despite the rather short instruction time of the video modeling examples (10 min), thereby attesting to their efficiency. Our video modeling examples did not only demonstrate how to perform unconfounded experiments but also provided narrative explanations of the model's thought processes. Thereby, self-regulation instruction was integrated within scientific reasoning instruction. Teachers can easily create video modeling examples and use them in science education. Consequently, teachers are increasingly using instructional videos in online platforms like YouTube or KhanAcademy. However, instructional videos are focused on teaching content, whereas we provided evidence that video modeling examples can also be used to convey specific learning strategies like scientific reasoning.

Limitations and Future Directions

Several limitations to our findings should be considered. First, since the scientific reasoning pre- and posttest items were identical and participants simply repeated their answers from pre- to posttest, we could not assess the effectiveness of the intervention on scientific reasoning ability reliably. Therefore, the effects of video modeling examples and metacognitive prompts on scientific reasoning ability should be investigated in future research. Nevertheless, the present study showed that video modeling examples improved scientific reasoning products and processes, which one might argue are more meaningful measures than declarative scientific reasoning knowledge. Second, due to technical difficulties with the audio recording, the process data analyses were conducted using a smaller sample size. A replication of the process analyses with a larger dataset would be beneficial for confirming the effectiveness of video modeling examples on the process level.

Third, the intervention effects were only partly visible in the transfer task, in which only a better hypothesis quality was achieved. This could indicate that the benefits of video modeling examples are not robust enough to attain transferable knowledge. Conceptual knowledge for the transfer task was lower than the training task; therefore, applying the learned processes to a more complex task could have been challenging for learners. Last, only hypothesis quality was explicitly modeled

in the videos. Future research should test these different explanations for the lack of transfer regarding argumentation quality, for example, with a longer delay between the training and the transfer tasks.

Furthermore, no benefits of the metacognitive prompts were observed on the product or process level. One explanation is that participants in the video modeling conditions were already self-regulating sufficiently in the training task, as identified by the process data. This finding suggests that video modeling examples were sufficient for fostering self-regulation in the training task. Therefore, providing the metacognitive prompts in the training task might have been unnecessary. One direction for future research would be to provide the metacognitive prompts in the transfer task to further support self-regulation after the intervention took place. Second, prior research on supporting monitoring and control has suggested that active generative tasks during (Mazzoni & Nelson, 1995) or after learning (Schleinschok et al., 2017; van Loon et al., 2014) can improve monitoring accuracy. Furthermore, providing metacognitive judgements was only effective for learning after retrieval (Ariel et al., 2021). Therefore, generating a written response to the prompt or retrieving information from the video modeling examples before the prompts might increase their effectiveness.

Conclusions

The present study provided evidence on the effectiveness of video modeling examples for improving scientific reasoning process and products. Watching video modeling examples improved hypothesis and argumentation quality in the training task and hypothesis quality in the transfer task. Students who watched video modeling examples were also self-regulating more during scientific reasoning activities, as indicated by the process analyses. Thus, an integrated instruction of self-regulation and scientific reasoning resulted in a conjoint use of these processes. Therefore, the present study provided evidence for the effectiveness of video-modeling examples both at the product and process level using fine-grained analyses regarding the co-occurrence and sequence of scientific reasoning and self-regulation processes. Our findings are applicable outside the context of science education as video modeling examples can be used to support other task-specific and self-regulatory processes in an integrated manner. Likewise, our methodological approach combining process and product measures, can also be applied to other questions in educational psychology.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10648-021-09652-3>.

Funding Open Access funding enabled and organized by Projekt DEAL. This research was supported as part of the LEAD Graduate School & Research Network [GSC1028], which was funded within the framework of the Excellence Initiative of the German federal and state governments.

Declarations

Conflict of Interest The experimental design and the hypotheses of this study were preregistered in As-Predicted (<https://aspredicted.org/vs43g.pdf>) prior to data collection. The authors declare no competing

interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abd-El-Khalick, F., BouJaoude, S., Duschl, R., Lederman, N. G., Mamlok-Naaman, R., Hofstein, A., ... Tuan, H. (2004). Inquiry in science education: International perspectives. *Science Education*, 88(3), 397–419. <https://doi.org/10.1002/sce.10118>
- Ariel, R., Karpicke, J. D., Witherby, A. E., & Tauber, S. K. (2021). Do judgments of learning directly enhance learning of educational materials? *Educational Psychology Review*, 33, 693–712. <https://doi.org/10.1007/s10648-020-09556-8>
- Azevedo, R. (2005). Using hypermedia as a metacognitive tool for enhancing student learning? The role of self-regulated learning. *Educational Psychologist*, 40(4), 199–209. https://doi.org/10.1207/s15326985ep4004_2
- Azevedo, R., & Hadwin, A. F. (2005). Scaffolding self-regulated learning and metacognition – Implications for the design of computer-based scaffolds. *Instructional Science*, 33(5–6), 367–379. <https://doi.org/10.1007/s11251-005-1272-9>
- Azevedo, R., Moos, D. C., Johnson, A. M., & Chauncey, A. D. (2010). Measuring cognitive and metacognitive regulatory processes during hypermedia learning: Issues and challenges. *Educational Psychologist*, 45(4), 210–223. <https://doi.org/10.1080/00461520.2010.515934>
- Azevedo, R., Martin, S. A., Taub, M., Mudrick, N. V., Millar, G. C., & Grafsgaard, J. F. (2016). Are pedagogical agents' external regulation effective in fostering learning with intelligent tutoring systems? In *International Conference on Intelligent Tutoring Systems* (pp. 197–207). Springer, Cham.
- Baars, M., van Gog, T., de Bruin, A., & Paas, F. G. W. C. (2014). Effects of problem solving after worked example study on primary school children's monitoring accuracy. *Applied Cognitive Psychology*, 28, 382–391. <https://doi.org/10.1002/acp.3008>
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs: Prentice Hall.
- Bannert, M. (2009). Supporting self-regulated hypermedia learning through prompts: A discussion. *Zeitschrift Für Pädagogische Psychologie*, 23(2), 139–145. <https://doi.org/10.1024/1010-0652.23.2.139>
- Bannert, M., Reimann, P., & Sonnenberg, C. (2014). Process mining techniques for analyzing patterns and strategies in students' self-regulated learning. *Metacognition and Learning*, 9(2), 161–185. <https://doi.org/10.1007/s11409-013-9107-6>
- Bannert, M., Sonnenberg, C., Mengelkamp, C., & Pieger, E. (2015). Short- and long-term effects of students' self-directed metacognitive prompts on navigation behavior and learning performance. *Computers in Human Behavior*, 52, 293–306. <https://doi.org/10.1016/j.chb.2015.05.038>
- Belland, B. R., Walker, A. E., Olsen, M. W., & Leary, H. (2015). A pilot meta-analysis of computer-based scaffolding in STEM education. *Journal of Educational Technology & Society*, 18(1), 183.
- Boekaerts, M. (1999). Self-regulated learning: Where we are today. *International Journal of Educational Research*, 31(6), 445–457. [https://doi.org/10.1016/s0883-0355\(99\)00014-2](https://doi.org/10.1016/s0883-0355(99)00014-2)
- Bolt, A., van der Aalst, W. M. P., & de Leoni, M. (2017). Finding process variants in event logs. In *OTM Confederated International Conferences "On the Move to Meaningful Internet Systems"* (pp. 45–52). Springer.

- Bradley, E. H., Curry, L. A., & Devers, K. J. (2007). Qualitative data analysis for health services research: Developing taxonomy, themes, and theory. *Health Services Research*, 42(4), 1758–1772. <https://doi.org/10.1111/j.1475-6773.2006.00684.x>
- Bussey, T. A., Tunnicliffe, J., Loftus, G. R., & Loftus, E. F. (2000). Accounts of the confidence-accuracy relation in recognition memory. *Psychonomic Bulletin & Review*, 7(1), 26–48. <https://doi.org/10.3758/bf03210724>
- Chen, Z., & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development*, 70(5), 1098–1120. <https://doi.org/10.1111/1467-8624.00081>
- Chi, M. T., & 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>
- Chin, C., & Brown, D. E. (2000). Learning in science: A comparison of deep and surface approaches. *Journal of Research in Science Teaching*, 37(2), 109–138. [https://doi.org/10.1002/\(sici\)1098-2736\(200002\)37:2%3c109::aid-tea3%3e3.0.co;2-7](https://doi.org/10.1002/(sici)1098-2736(200002)37:2%3c109::aid-tea3%3e3.0.co;2-7)
- Clark, D., & Linn, M. C. (2009). Designing for knowledge integration: The impact of instructional time. *Journal of Education*, 189(1–2), 139–158. <https://doi.org/10.1177/0022057409189001-210>
- Cooper, G., & Sweller, J. (1987). Effects of schema acquisition and rule automation on mathematical problem-solving transfer. *Journal of Educational Psychology*, 79(4), 347–362. <https://doi.org/10.1037/0022-0666.79.4.347>
- Csanadi, A., Eagan, B., Kollar, I., Shaffer, D. W., & Fischer, F. (2018). When coding-and-counting is not enough: Using epistemic network analysis (ENA) to analyze verbal data in CSDL research. *International Journal of Computer-Supported Collaborative Learning*, 13(4), 419–438. <https://doi.org/10.1007/s11412-018-9292-z>
- Davis, E. A. (2000). Scaffolding students' knowledge integration: Prompts for reflection in KIE. *International Journal of Science Education*, 22(8), 819–837. <https://doi.org/10.1080/095006900412293>
- de Jong, T. (2006). Computer simulations-technological advances in inquiry learning. *Science*, 312, 532–533. <https://doi.org/10.1126/science.1127750>
- de Jong, T., & Lazonder, A. W. (2014). The guided discovery principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 371–390). Cambridge University Press.
- de Jong, T., & van Joolingen, W. R. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research*, 68(2), 179–201. <https://doi.org/10.3102/00346543068002179>
- Dori, Y. J., Avargil, S., Kohen, Z., & Saar, L. (2018). Context-based learning and metacognitive prompts for enhancing scientific text comprehension. *International Journal of Science Education*, 40(10), 1198–1220. <https://doi.org/10.1080/09500693.2018.1470351>
- Engelmann, K., Neuhaus, B. J., & Fischer, F. (2016). Fostering scientific reasoning in education—meta-analytic evidence from intervention studies. *Educational Research and Evaluation*, 22(5–6), 333–349. <https://doi.org/10.1080/13803611.2016.1240089>
- Engelmann, K., & Bannert, M. (2019). Analyzing temporal data for understanding the learning process induced by metacognitive prompts. *Learning and Instruction*. <https://doi.org/10.1016/j.learninstruc.2019.05.002>
- 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, 175–191. <https://doi.org/10.3758/bf03193146>
- Fischer, F., Kollar, I., Ufer, S., Sodian, B., Hussmann, H., Pekrun, R., ... Eberle, J. (2014). Scientific reasoning and argumentation: Advancing an interdisciplinary research agenda in education. *Frontline Learning Research*, 2(3), 28–45. <https://doi.org/10.14786/flr.v2i2.96>
- Furtak, E. M., Seidel, T., Iverson, H., & Briggs, D. C. (2012). Experimental and quasi-experimental studies of inquiry-based science teaching. *Review of Educational Research*, 82, 300–329. <https://doi.org/10.3102/0034654312457206>
- Fyfe, E. R., & Rittle-Johnson, B. (2016). Feedback both helps and hinders learning: The causal role of prior knowledge. *Journal of Educational Psychology*, 108(1), 82–97. <https://doi.org/10.1037/edu0000053>
- Grüß-Niehaus, T. (2010). *Zum Verständnis des Löslichkeitskonzeptes im Chemieunterricht - der Effekt von Methoden progressiver und kollaborativer Reflexion* [Comprehension of the dissolution concept in chemistry education: The effect of methods of progressive and collaborative reflection]. Hannover, Germany: Gottfried Wilhelm Leibniz Universität.

- Hadwin, A. F., & Winne, P. H. (2001). CoNoteS2: A software tool for promoting self-regulation. *Educational Research and Evaluation*, 7, 313–334. <https://doi.org/10.1076/edre.7.2.313.3868>
- Hartmann, C., Gog, T., & Rummel, N. (2020). Do examples of failure effectively prepare students for learning from subsequent instruction? *Applied Cognitive Psychology*, 34(4), 879–889. <https://doi.org/10.1002/acp.3651>
- Hartmann, S., Upmeyer zu Belzen, A., Krüger, D., & Pant, H. A. (2015). Scientific reasoning in higher education: Constructing and evaluating the criterion-related validity of an assessment of preservice science teachers' competencies. *Zeitschrift für Psychologie*, 223(1), 47–53. <https://doi.org/10.1027/2151-2604/a000199>
- Hetmanek, A., Engelmann, K., Opitz, A., & Fischer, F. (2018). Beyond intelligence and domain knowledge. In F. Fischer, C. A. Chinn, K. Engelmann, & J. Osborne (Eds.), *Scientific reasoning and argumentation* (pp. 203–226). New York, NY: Routledge. <https://doi.org/10.4324/9780203731826-12>
- Hidi, S. (1990). Interest and its contribution as a mental resource for learning. *Review of Educational Research*, 60(4), 549–571. <https://doi.org/10.3102/00346543060004549>
- Hidi, S., & Renninger, A. (2006). The four-phase model of interest development. *Educational Psychologist*, 41(2), 111–127. https://doi.org/10.1207/s15326985ep4102_4
- Hidi, S., & Ainley, M. (2008). Interest and self-regulation: Relationships between two variables that influence learning. In D. H. Schunk & B. J. Zimmerman (Eds.), *Motivation and self-regulated learning: Theory, research and applications* (pp. 77–109). Mahwah, NJ: Lawrence Erlbaum and Associates
- 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>
- Huang, C. (2011). Self-concept and academic achievement: A meta-analysis of longitudinal relations. *Journal of School Psychology*, 49, 505–528. <https://doi.org/10.1016/j.jsp.2011.07.001>
- Kant, J. M., Scheiter, K., & Oschatz, K. (2017). How to sequence video modeling examples and inquiry tasks to foster scientific reasoning. *Learning and Instruction*, 52, 46–58. <https://doi.org/10.1016/j.learninstruc.2017.04.005>
- Kim, H. J., & Pedersen, S. (2011). Advancing young adolescents' hypothesis-development performance in a computer-supported and problem-based learning environment. *Computers & Education*, 57(2), 1780–1789. <https://doi.org/10.1016/j.compedu.2011.03.014>
- Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, 12(1), 1–48. [https://doi.org/10.1016/0364-0213\(88\)90007-9](https://doi.org/10.1016/0364-0213(88)90007-9)
- Koslowski, B. (2012). Scientific reasoning: Explanation, confirmation bias, and scientific practice. In G. J. Feist & M. E. Gorman (Eds.), *Handbook of the psychology of science* (pp. 151–192). Springer.
- Kostons, D., Van Gog, T., & Paas, F. (2012). Training self-assessment and task-selection skills: A cognitive approach to improving self-regulated learning. *Learning and Instruction*, 22, 121–132. <https://doi.org/10.1016/j.learninstruc.2011.08.004>
- Krüger, D., Hartmann, S., Nordmeier, V., & Upmeyer zu Belzen, A. (2020). Measuring scientific reasoning competencies – Multiple aspects of validity. In O. Zlatkin-Troitschanskaia, H. Pant, M. Toepper, & C. Lautenbach (Eds.), *Student Learning in German Higher Education* (pp. 261–280). Springer.
- Kuhn, D., Black, J., Keselman, A., & Kaplan, D. (2000). The development of cognitive skills to support inquiry learning. *Cognition and Instruction*, 18(4), 495–523. https://doi.org/10.1207/s1532690xc1804_3
- Kuhn, D. (1991). *The skills of argument*. Cambridge University Press.
- Lai, C.-L., Hwang, G.-J., & Tu, Y.-H. (2018). The effects of computer-supported self-regulation in science inquiry on learning outcomes, learning processes, and self-efficacy. *Educational Technology Research and Development*, 66(4), 863–892. <https://doi.org/10.1007/s11423-018-9585-y>
- Lazonder, A. W., & Harmsen, R. (2016). Meta-analysis of inquiry-based learning: Effects of guidance. *Review of Educational Research*, 86(3), 681–718. <https://doi.org/10.3102/0034654315627366>
- Lin, X. (2001). Designing metacognitive activities. *Educational Technology Research and Development*, 49(2), 23–40. <https://doi.org/10.1007/bf02504926>
- Mäeots, M., Siiman, L., Kori, K., Eelmeets, M., Pedaste, M., & Anjewierden, A. (2016). The role of a reflection tool in enhancing students' reflection. *10th Annual International Technology, Education and Development Conference (INTED 2016)* (pp. 1892–1900).

- Manlove, S., Lazonder, A. W., & de Jong, T. (2007). Software scaffolds to promote regulation during scientific inquiry learning. *Metacognition and Learning*, 2, 141–155. <https://doi.org/10.1007/s11409-007-9012-y>
- Manlove, S., Lazonder, A. W., & de Jong, T. (2009). Trends and issues of regulative support use during inquiry learning: Patterns from three studies. *Computers in Human Behavior*, 25(4), 795–803. <https://doi.org/10.1016/j.chb.2008.07.010>
- Marquart, C. L., Hinojosa, C., Swiecki, Z., Eagan, B., & Shaffer, D. W. (2018). Epistemic Network Analysis (Version 1.7.0) [Computer software]. Available from <http://app.epistemicnetwork.org>
- Marsh, H. W., & Martin, A. J. (2011). Academic self-concept and academic achievement: Relations and causal ordering. *British Journal of Educational Psychology*, 81, 59–77. <https://doi.org/10.1348/000709910X503501>
- Mazzoni, G., & Nelson, T. O. (1995). Judgments of learning are affected by the kind of encoding in ways that cannot be attributed to the level of recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 1263–1274. <https://doi.org/10.1037/0278-7393.21.5.1263>
- McLaren, B. M., Lim, S., & Koedinger, K. R. (2008). When and how often should worked examples be given to students? New results and a summary of the current state of research. In B. C. Love, K. McRae, & V.M. Sloutsky (Eds.), *Proceedings of the 30th annual conference of the cognitive science society* (pp. 2176–2181). Austin: Cognitive Science Society.
- McNeill, K. L. (2011). Elementary students' views of explanation, argumentation, and evidence, and their abilities to construct arguments over the school year. *Journal of Research in Science Teaching*, 48(7), 793–823. <https://doi.org/10.1002/tea.20430>
- McNeill, K. L., Lizotte, D. J., Krajcik, J., & Marx, R. W. (2006). Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. *Journal of the Learning Sciences*, 15(2), 153–191. https://doi.org/10.1207/s15327809jls1502_1
- Mulder, Y. G., Lazonder, A. W., & de Jong, T. (2014). Using heuristic worked examples to promote inquiry-based learning. *Learning and Instruction*, 29, 56–64. <https://doi.org/10.1016/j.learninstruc.2013.08.001>
- Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 125–173). Academic Press.
- OECD. (2013). *Trends shaping education 2013*. Paris, France: Author. https://doi.org/10.1787/trends_edu-2013-en
- Omarchevska, Y., Lachner, A., Richter, J., & Scheiter, K. (2021). It takes two to tango: How scientific reasoning and self-regulation processes impact argumentation quality. *Journal of the Learning Sciences*, 1–41. <https://doi.org/10.1080/10508406.2021.1966633>
- Ommundsen, Y., Haugen, R., & Lund, T. (2005). Academic self-concept, implicit theories of ability, and self-regulation strategies. *Scandinavian Journal of Educational Research*, 49(5), 461–474. <https://doi.org/10.1080/00313830500267838>
- Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching*, 41(10), 994–1020. <https://doi.org/10.1002/tea.20035>
- Panadero, E., & Alonso-Tapia, J. (2014). How do students self-regulate? Review of Zimmerman's cyclical model of self-regulated learning. *Anales De Psicología*, 30(2), 450–462. <https://doi.org/10.6018/analesps.30.2.167221>
- Pedaste, M., Mäeots, M., Siiman, L. A., de Jong, T., van Riesen, S. A. N., Kamp, E. T., & Tsourlidaki, E. (2015). Phases of inquiry-based learning: Definitions and the inquiry cycle. *Educational Research Review*, 14, 47–61. <https://doi.org/10.1016/j.edurev.2015.02.003>
- Pintrich, P. R. (1999). The role of motivation in promoting and sustaining self-regulated learning. *International Journal of Educational Research*, 31(6), 459–470. [https://doi.org/10.1016/s0883-0355\(99\)00015-4](https://doi.org/10.1016/s0883-0355(99)00015-4)
- Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.) *Handbook of self-regulation* (pp. 451–502). Academic Press. <https://doi.org/10.1016/b978-012109890-2/50043-3>
- Quintana, C., Reiser, B. J., Davis, E. A., Krajcik, J., Fretz, E., Duncan, R. G., & Soloway, E. (2004). A scaffolding design framework for software to support science inquiry. *The Journal of the Learning Sciences*, 13, 337–387. https://doi.org/10.1207/s15327809jls1303_4
- Quintana, C., Zhang, M., & Krajcik, J. (2005). A framework for supporting metacognitive aspects of online inquiry through software-based scaffolding. *Educational Psychologist*, 40(4), 235–244. https://doi.org/10.1207/s15326985ep4004_5

- Raaijmakers, S. F., Baars, M., Paas, F., van Merriënboer, J. J., & Van Gog, T. (2018a). Training self-assessment and task-selection skills to foster self-regulated learning: Do trained skills transfer across domains? *Applied Cognitive Psychology*, 32(2), 270–277. <https://doi.org/10.1002/acp.3392>
- Raaijmakers, S. F., Baars, M., Schaap, L., Paas, F., Van Merriënboer, J., & Van Gog, T. (2018b). Training self-regulated learning skills with video modeling examples: Do task-selection skills transfer? *Instructional Science*, 46(2), 273–290. <https://doi.org/10.1007/s11251-017-9434-0>
- Reid, D. J., Zhang, J., & Chen, Q. (2003). Supporting scientific discovery learning in a simulation environment. *Journal of Computer Assisted Learning*, 19(1), 9–20. <https://doi.org/10.1046/j.0266-4909.2003.00002.x>
- Reid, A. J., Morrison, G. R., & Bol, L. (2017). Knowing what you know: Improving metacomprehension and calibration accuracy in digital text. *Educational Technology Research & Development*, 65(1), 29–45. <https://doi.org/10.1007/s11423-016-9454-5>
- Reimann, P. (2009). Time is precious: Variable- and event-centred approaches to process analysis in CSCL research. *International Journal of Computer-Supported Collaborative Learning*, 4(3), 239–257. <https://doi.org/10.1007/s11412-009-9070-z>
- Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cognitive Science*, 38, 1–37. <https://doi.org/10.1111/cogs.12086>
- Renkl, A. (2018). Scientific reasoning and argumentation: Is there an over-emphasis on discipline specificity? In F. Fischer, C. A. Chinn, K. Engelmann, & J. Osborne (Eds.), *Scientific reasoning and argumentation: The roles of domain-specific and domain-general knowledge* (pp. 194–200). New York, NY: Routledge. <https://doi.org/10.4324/9780203731826-11>
- Renninger, K. A., & Hidi, S. E. (2019). Interest development and learning. In K. A. Renninger & S. E. Hidi (Eds.), *The Cambridge handbook of motivation and learning* (pp. 265–290). Cambridge University Press.
- Roll, I., & Winne, P. H. (2015). Understanding, evaluating, and supporting self-regulated learning using learning analytics. *Journal of Learning Analytics*, 2, 7–12. <https://doi.org/10.18608/jla.2015.21.2>
- Romero, C., Ventura, S., Pechenizkiy, M., & Baker, R. S. (2010). *Handbook of educational data mining*. New York: Taylor & Francis.
- Schanze, S. (2002). *Wissenserwerb mithilfe der internetbasierten Lernumgebung ChemNet: Eine empirische Untersuchung zum Lernen mit linearen und vernetzten Hypertexten* [Knowledge acquisition by means of the online learning environment ChemNet: An empirical study of linear and cross-linked hypertexts]. Kiel, Germany: Christian-Albrechts-Universität zu Kiel.
- Scheiter, K., & Gerjets, P. (2007). Learner control in hypermedia environments. *Educational Psychology Review*, 19(3), 285–307. <https://doi.org/10.1007/s10648-007-9046-3>
- Schleinschok, K., Eitel, A., & Scheiter, K. (2017). Do drawing tasks improve monitoring and control during learning from text? *Learning and Instruction*, 51, 10–25. <https://doi.org/10.1016/j.learninstruc.2017.02.002>
- Shunk, D. H., & Zimmerman, B. J. (Eds.). (1998). *Self-regulated learning: From teaching to self-reflective practice*. Guilford Press.
- Shaffer, D. W., Collier, W., & Ruis, A. R. (2016). A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *Journal of Learning Analytics*, 3(3), 9–45. <https://doi.org/10.18608/jla.2016.33.3>
- Shunk, D. H., & Zimmerman, B. J. (Eds.). (1998). *Self-regulated learning: From teaching to self-reflective practice*. Guilford Press.
- Smit, K., de Brabander, C. J., Boekaerts, M., & Martens, R. L. (2017). The self-regulation of motivation: Motivational strategies as mediator between motivational beliefs and engagement for learning. *International Journal of Educational Research*, 82, 124–134. <https://doi.org/10.1016/j.ijer.2017.01.006>
- Sonnenberg, C., & Bannert, M. (2015). Discovering the effects of metacognitive prompts on the sequential structure of SRL-processes using process mining techniques. *Journal of Learning Analytics*, 2(1), 72–100. <https://doi.org/10.18608/jla.2015.21.5>
- Sonnenberg, C., & Bannert, M. (2019). Using process mining to examine the sustainability of instructional support: How stable are the effects of metacognitive prompting on self-regulatory behavior? *Computers in Human Behavior*, 96, 259–279. <https://doi.org/10.1016/j.chb.2018.06.003>
- Stadler, M., Sailer, M., & Fischer, F. (2021). Knowledge as a formative construct: A good alpha is not always better. *New Ideas in Psychology*, 60, 100832. <https://doi.org/10.1016/j.newideapsych.2020.100832>

- Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York: Springer.
- Trautwein, U., & Möller, J. (2016). Self-concept: Determinants and consequences of academic self-concept in school contexts. In A. A. Lipnevich, F. Preckel, & R. D. Roberts (Eds.), *Psychosocial skills and school systems in the 21st century: Theory, research, and practice* (pp. 187–214). Springer. https://doi.org/10.1007/978-3-319-28606-8_8
- Van der Aalst, W. (2016). *Process mining – Data science in action*. Berlin: Springer.
- Van Gog, T. (2011). Effects of identical example-problem and problem-example pairs on learning. *Computers & Education*, 57, 1775–1779. <https://doi.org/10.1016/j.compedu.2011.03.019>
- Van Gog, T., & Rummel, N. (2010). Example-based learning: Integrating cognitive and social-cognitive research perspectives. *Educational Psychology Review*, 22, 155–174. <https://doi.org/10.1007/s10648-010-9134-7>
- Van Gog, T., Jarodzka, H., Scheiter, K., Gerjets, P., & Paas, F. (2009). Attention guidance during example study via the model's eye movements. *Computers in Human Behavior*, 25, 785–791. <https://doi.org/10.1016/j.chb.2009.02.007>
- Van Joolingen, W. R., & Zacharia, Z. C. (2009). Developments in inquiry learning. In N. Balacheff, S. Ludvigsen, T. de Jong, & S. Barnes (Eds.), *Technology-enhanced learning* (pp. 21–37). Dordrecht, The Netherlands: Springer. https://doi.org/10.1007/978-1-4020-9827-7_2
- van Loon, M. H., de Bruin, A. B. H., van Gog, T., van Merriënboer, J. J. G., & Dunlosky, J. (2014). Can students evaluate their understanding of cause-and-effect relations? The effects of diagram completion on monitoring accuracy. *Acta Psychologica*, 151, 143–154. <https://doi.org/10.1016/j.actpsy.2014.06.007>
- Verbeek, H. M. W., Buijs, J. C. A. M., Van Dongen, B. F., & van der Aalst, W. M. (2010). Prom 6: The process mining toolkit. *Proceedings of BPM Demonstration Track*, 615, 34–39.
- Weijters, A. J. M. M., van der Aalst, W. M., & De Medeiros, A. A. (2006). Process mining with the heuristics miner-algorithm. *Technische Universiteit Eindhoven. Tech. Rep. WP*, 166, 1–34.
- White, B. Y., & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction*, 16(1), 3–118. https://doi.org/10.1207/s1532690xc11601_2
- White, B. Y., Frederiksen, J. R., & Collins, A. (2009). The interplay of scientific inquiry and metacognition. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 175–205). Routledge.
- Wijnia, L., & Baars, M. (2021). The role of motivational profiles in learning problem-solving and self-assessment skills with video modeling examples. *Instructional Science*, 49(1), 67–107. <https://doi.org/10.1007/s11251-020-09531-4>
- Wilde, M., Bätz, K., Kovaleva, A., & Urhahne, D. (2009). Überprüfung einer Kurzskaala intrinsischer Motivation (KIM) [Review of a short scale for intrinsic motivation]. *Zeitschrift für Didaktik der Naturwissenschaften* [Journal for the didactics of science], 15, 31–45.
- Zacharia, Z. C., Manoli, C., Xenofontos, N., de Jong, T., Pedaste, M., van Riesen, S. A. N., ... Tsourlidaki, E. (2015). Identifying potential types of guidance for supporting student inquiry when using virtual and remote labs in science: a literature review. *Educational Technology Research and Development*, 63(2), 257–302. <https://doi.org/10.1007/s11423-015-9370-0>
- Zheng, L. (2016). The effectiveness of self-regulated learning scaffolds on academic performance in computer-based learning environments: A meta-analysis. *Asia Pacific Education Review*, 17(2), 187–202.
- Zimmerman, B. J. (2013). From cognitive modeling to self-regulation: A social cognitive career path. *Educational Psychologist*, 48(3), 135–147. <https://doi.org/10.1080/00461520.2013.794676>
- Zimmerman, B. J., & Moylan, A. R. (2009). Self-regulation: Where metacognition and motivation intersect. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of Metacognition in Education* (pp. 299–315). Routledge.