The Interplay of Scientific Reasoning and Self-Regulation in Inquiry Learning: A Temporal Process-Oriented Analysis

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
Yoana Omarchevska
aus Sofia/Bulgarien

Tübingen
2022

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	19.07.2022
Dekan:	Prof. Dr. Thilo Stehle
1. Berichterstatter/-in:	Prof. Dr. Katharina Scheiter
2. Berichterstatter/-in:	Prof. Dr. Caterina Gawrilow
3. Berichterstatter/-in:	Prof. Dr. Frank Fischer

“Thinking scientifically entails adopting a scientific attitude, an appreciation that evidence is important and the willingness to change views in light of new evidence. An individual may struggle to pick up scientific thinking without deliberate and thoughtfully crafted instruction designed to promote it. The disposition to think critically and analytically must be encouraged because our natural tendency is to make quick judgments and move on. Individuals may be unwilling or unable to do the heavy lifting scientific thinking requires.”

Gale Sinatra & Barbara Hofer, 2021

Acknowledgements

I would like to express my gratitude to everyone who supported me during my dissertation. First and foremost, I would like to thank Prof. Dr. Katharina Scheiter, for all the time and effort she invested in helping, advising, and supervising me. Thank you, Katharina, for all your support, feedback, ideas, and for finding the perfect balance between providing me with guidance and freedom. I would like to thank my second supervisor Prof. Dr. Andreas Lachner, for his ideas, advice, and quick and helpful feedback. I would like to thank Prof. Dr. Juliane Richter for supporting me in the early stages of my PhD and answering all my questions.

I would like to also thank Prof. Dr. Caterina Gawrilow for being my second reviewer. I would like to thank the Leibniz-Institut für Wissensmedien for providing me with a supportive working environment and the LEAD Graduate School & Network for the networking opportunities and inspiring retreats.

Next, I would like to thank my colleagues in the AG Multiple Representations for their support and help with navigating the German language and bureaucracy. Thank you, Salome Wörner, for being there for me through all the ups and downs, for your friendship, for the care package you made that helped me write my dissertation, and for all the nice moments we shared together. Thank you, Emely Hoch, for all your support and advice in professional and personal context, and the friendship that followed.

Finally, I would like to say a special thank you to my family – to my father who inspired me to pursue a career in academia in the first place; to my mother who has always been my greatest supporter; to my sister who has always been there to give me advice and emotional support. Just as importantly, I would like to say a very special thank you to Yannick Burkart for encouraging me, believing in me, supporting me emotionally, and reminding me of what is important in life.

Table of Contents

TABLE OF CONTENTS

Summary

The improvement of scientific reasoning and argumentation has become a central aim of science education, which resulted in an increased implementation of active learner-centered pedagogies like inquiry learning (KMK, 2020; NASEM, 2021). Since inquiry learning relies on students' active engagement and participation (Freeman et al., 2014), it is important to consider factors related to its effective implementation. This dissertation has three conceptual aims related to investigating 1) the combined influence of students' cognitive and motivational characteristics on their experimentation skills and conceptual understanding, 2) the interplay of self-regulation and scientific reasoning processes and argumentation quality, and 3) the effectiveness of integrated scaffolding of self-regulation and scientific reasoning for students' learning processes and learning outcomes in the context of inquiry learning. Three studies were conducted to investigate these conceptual aims.

Study 1 examined how cognitive and motivational characteristics conjointly relate to students' experimentation skills and conceptual understanding in a guided inquiry learning science lesson. Secondary school students from six classes (N = 110, M_{age} = 12.07 years) attended a lesson and solved a structured inquiry task using a simulation on the topic of photosynthesis. Students' experimentation skills comprised the proportion of controlled experiments they conducted. Argumentation quality was measured from students' answers to the research question of the inquiry task. Conceptual understanding comprised students' written argumentation quality and posttest scores. Results revealed three distinct clusters, based on students' prior knowledge, academic self-concept, and interest: Underestimating, Struggling, and Strong. Underestimating students had high prior knowledge, but low interest and self-concept. Struggling students scored below average on all three variables, whereas Strong students scored above average on all variables. Struggling students performed significantly higher proportion of controlled experiments than Strong students while Underestimating and Strong significantly outperformed them in the conceptual knowledge posttest. There were no differences between the clusters in argumentation quality. A mediation analysis found no significant mediation of experimentation skills for the relationship between students' individual prerequisites and conceptual understanding, suggesting that the proportion of controlled experiments might be a too limited measure of students' experimentation skills.

Study 2 investigated the interplay of scientific reasoning and self-regulation processes in students with high and low argumentation quality. University students (N = 30, M_{age} = 23.33 years) solved an inquiry task on the topic of population genetics that asked them to test an assumption made by a fictitious person using a computer simulation. At the end of the task, they wrote an explanation answering the original research question, from which argumentation quality was coded. Fine-grained process data, obtained from screen recordings and think aloud protocols during the inquiry task, were used to explore the co-occurrences of scientific reasoning and self-regulation processes in relation to argumentation quality. A comparison between students with high and low argumentation quality using epistemic network analysis (ENA) showed that high argumentation quality students performed different scientific reasoning processes conjointly (and correctly) more frequently than low argumentation quality students. Importantly, more frequent co-occurrences of self-regulation processes (e.g., monitoring) and scientific reasoning activities were observed in the high argumentation quality group. These findings suggest that instruction, which integrates self-regulation and scientific reasoning, might be beneficial for improving scientific reasoning and argumentation.

Study 3 investigated the effectiveness of an integrated instruction of self-regulation and scientific reasoning using video modeling examples (VM) and metacognitive prompts (P), designed based on the results of Study 2. In two inquiry tasks, the effects of watching VM examples combined with prompts (VMP) to watching VM examples only, and to unguided inquiry (control) were compared. The intervention was evaluated with respect to university students' (N = 127, M_{age} = 24.3 years) scientific reasoning ability, hypothesis and argumentation quality, and scientific reasoning and self-regulation processes. Results showed that participants in the VM conditions had higher hypothesis and argumentation quality in the training task and higher hypothesis quality in the transfer task than the control group. No differences between the two VM conditions were found, suggesting no additional benefit of the prompts. Students' conjoint and sequential use of self-regulation and scientific reasoning processes were modeled using ENA and process mining. Stronger co-occurrences between scientific reasoning and self-regulation processes in the two VM conditions compared to the control condition were found using ENA, whereas process mining additionally revealed specific sequences of these processes. These findings indicate that an instruction that integrated self-regulation and scientific reasoning resulted in more conjoint use of scientific reasoning and self-regulation processes, and higher hypothesis and argumentation quality. Therefore, the integrated instruction of self-regulation and scientific reasoning improved students' learning processes and learning outcomes. Findings from this dissertation indicate that students' prior cognitive and motivational prerequisites, self-regulation processes, and scientific reasoning processes need to be considered in addition to learning outcomes during inquiry learning.

Zusammenfassung

Wissenschaftliches Denken und Argumentieren zu verbessern ist ein zentrales Ziel naturwissenschaftlichen Unterrichts, weshalb aktive lernerzentrierte Unterrichtsformen wie beispielsweise das forschende Lernen heutzutage vermehrt eingesetzt werden (KMK, 2020; NASEM, 2021). Da forschendes Lernen auf dem aktiven Engagement und der Beteiligung der Lernenden beruht (Freeman et al., 2014), ist es wichtig, Faktoren zu berücksichtigen, die mit der effektiven Umsetzung dieser Lernform zusammenhängen. In dieser Dissertation werden drei konzeptionelle Ziele verfolgt: 1) der kombinierte Einfluss der kognitiven und motivationalen Merkmale von Schülerinnen und Schülern auf ihre Experimentierfähigkeiten und ihr konzeptionelles Verständnis, 2) das Zusammenspiel von Selbstregulation und wissenschaftlichen Argumentationsprozessen sowie die Argumentationsqualität und 3) die Wirksamkeit von integrierten Hilfestellungen zu Selbstregulation und wissenschaftlichem Denken für die Lernprozesse und Lernergebnisse der Schülerinnen und Schüler im Kontext des forschenden Lernens. Drei Studien wurden durchgeführt, um diese konzeptionellen Ziele zu untersuchen.

In Studie 1 wurde untersucht, wie kognitive und motivationale Fähigkeiten gemeinsam mit den Experimentierfähigkeiten und dem konzeptionellen Verständnis der Schülerinnen und Schüler im angeleiteten forschenden Naturwissenschaftsunterricht zusammenhängen. Sekundarschülerinnen und -schüler aus sechs Klassen (N = 110, M_{Alter} = 12.07 Jahre) sollten im Rahmen einer Biologie-Unterrichtsstunde mit Hilfe einer Simulation zum Thema Photosynthese eine angeleitete Experimentieraufgabe lösen. Die Experimentierfähigkeiten der Schülerinnen und Schüler wurden durch den Anteil der von ihnen kontrolliert durchgeführten Experimente erfasst. Die Argumentationsqualität der Schülerinnen und Schüler wurde anhand ihrer Antworten auf die Forschungsfrage zu der Experimentieraufgabe gemessen. Das konzeptionelle Verständnis wurde durch die Qualität der schriftlichen Argumentationen der Schülerinnen und Schüler und ihrer Ergebnisse im Posttest erfasst. Die Ergebnisse zeigten drei verschiedene Cluster von Schülerinnen und Schülern, basierend auf ihrem Vorwissen, ihrem akademischen Selbstkonzept und ihrem Interesse: sich selbst unterschätzende, schwache und starke Schülerinnen und Schüler. Sich selbst unterschätzende Schülerinnen und Schüler hatten ein hohes Vorwissen, aber ein geringes Interesse und Selbstkonzept. Die schwachen Schülerinnen und Schüler erzielten bei allen drei Variablen unterdurchschnittliche Ergebnisse, während die starken Schüler bei allen Variablen überdurchschnittliche Ergebnisse erzielten.

Die schwachen Schülerinnen und Schüler führten einen signifikant höheren Anteil an kontrolliert durchgeführten Experimenten durch als starke Schülerinnen und Schüler, während Schülerinnen und Schüler, die sich selbst unterschätzten, und starke Schülerinnen und Schüler beim Posttest zum konzeptuellen Verständnis signifikant besser abschnitten. Bei der Argumentationsqualität gab es keine Unterschiede zwischen den Clustern. Eine Mediationsanalyse zeigte keine signifikante Mediation der Experimentierfähigkeit auf die Beziehung zwischen den individuellen Voraussetzungen der Schülerinnen und Schüler und ihrem konzeptionellen Verständnis, was darauf hindeutet, dass der Anteil der kontrolliert durchgeführten Experimente nicht aussagekräftig genug für die Experimentierfähigkeit der Schülerinnen und Schüler sein könnte.

In Studie 2 wurde das Zusammenspiel von wissenschaftlichem Denken und Selbstregulationsprozessen bei Studierenden mit hoher und niedriger Argumentationsqualität untersucht. Universitätsstudierende (N = 30, M_{Alter} = 23.33 Jahre) lösten eine Experimentieraufgabe zum Thema Populationsgenetik, bei der sie eine von einer fiktiven Person gemachten Annahme mit Hilfe einer Computersimulation überprüfen sollten. Am Ende der Aufgabe schrieben sie eine Erklärung, die die ursprüngliche Forschungsfrage beantwortete, und die Qualität ihrer Argumentation in dieser Erklärung wurde kodiert. Detaillierte Prozessdaten, die aus Bildschirmaufzeichnungen und Protokollen des lauten Denkens während der Experimentieraufgabe gewonnen wurden, wurden verwendet, um das gemeinsame Auftreten von wissenschaftlichem Denken und Selbstregulationsprozessen in Bezug auf die Argumentationsqualität zu untersuchen. Ein Vergleich zwischen Studierenden mit hoher und niedriger Argumentationsqualität unter Verwendung der epistemischen Netzwerkanalyse (ENA) zeigte, dass Studierende mit hoher Argumentationsqualität verschiedene wissenschaftliche Argumentationsprozesse häufiger gemeinsam (und korrekt) durchführten als Studierende mit niedriger Argumentationsqualität. Besonders wichtig ist, dass in der Gruppe mit hoher Argumentationsqualität häufiger Selbstregulationsprozesse (z. B. Monitoring) und Aktivitäten des wissenschaftlichen Denkens zusammen auftraten. Diese Ergebnisse deuten darauf hin, dass Unterricht, der Selbstregulation und wissenschaftliches Denken integriert, für die Verbesserung des wissenschaftlichen Denkens und Argumentierens von Vorteil sein könnte.

In Studie 3 wurde die Wirksamkeit von Unterricht untersucht, der Selbstregulation und wissenschaftliches Denken mittels Videomodellierungsbeispielen (VM) und metakognitiven

Prompts (P) integrierte und der auf der Grundlage der Ergebnisse von Studie 2 konzipiert wurde. In zwei Experimentieraufgaben wurde der Effekt des Betrachtens von VM-Beispielen in Kombination mit Prompts (VMP) mit dem des Betrachtens von VM-Beispielen allein und mit dem von unangeleitetem Experimentieren (Kontrollgruppe) verglichen. Die Intervention wurde dahingehend bewertet, welche Fähigkeiten Universitätsstudierende (N = 127, M_{Alter} = 24.3 Jahre) zum wissenschaftlichen Denken, zum Aufstellen hochqualitativer Hypothesen und Argumente sowie zu Prozessen des wissenschaftlichen Denkens und der Selbstregulation aufwiesen. Die Ergebnisse zeigten, dass die Teilnehmerinnen und Teilnehmer der VM-Bedingungen eine höhere Hypothesen- und Argumentationsqualität in der Trainingsaufgabe und eine höhere Hypothesenqualität in der Transferaufgabe hatten als die Kontrollgruppe. Es wurden keine Unterschiede zwischen den beiden VM-Bedingungen festgestellt, was darauf hindeutet, dass die metakognitiven Prompts keinen zusätzlichen Nutzen hatten. Das gemeinsame Auftreten von Selbstregulations- und wissenschaftlichen Argumentationsprozessen bei Studierenden wurde mit ENA und Process Mining modelliert. Die ENA zeigte, dass in den beiden VM-Bedingungen im Vergleich zur Kontrollgruppe die Prozesse des wissenschaftlichen Denkens und der Selbstregulation vermehrt gemeinsam auftraten, während das Process Mining zusätzlich spezifische Sequenzen dieser Prozesse aufzeigte. Diese Ergebnisse deuten darauf hin, dass Unterricht, der Selbstregulation und wissenschaftliches Denken integriert, zu einem stärkeren gemeinsamen Auftreten von Prozessen des wissenschaftlichen Denkens und der Selbstregulation sowie zu einer höheren Hypothesen- und Argumentationsqualität führt. Der Selbstregulation und wissenschaftliches Denken integrierende Unterricht verbesserte also die Lernprozesse und Lernergebnisse der Studierenden. Die Ergebnisse dieser Dissertation deuten darauf hin, dass beim forschenden Lernen neben den Lernergebnissen auch die kognitiven und motivationalen Voraussetzungen, die Selbstregulationsprozesse und die Prozesse des wissenschaftlichen Denkens der Lernenden berücksichtigt werden müssen.

List of Publications and Contributions

The manuscripts of Study 2 and 3 have been published elsewhere. The manuscript of Study 1 is currently under review in the journal *Learning and Instruction*. This list presents all publications included in this dissertation and the proportional contribution of the doctoral candidate.

Manuscript of Study 1

Author	Author position	Scientific ideas %	Data generation %	Analysis & Interpretation %	Paper writing %
Yoana Omarchevska	First	65%	15%	90%	85%
Andreas Lachner	Second				
Leonie Jacob	Third				
Juliane Richter	Fourth				
Katharina Scheiter	Fifth				

Title of paper: How do students' cognitive and motivational characteristics relate to experimentation skills and conceptual understanding during scientific inquiry?

Status in publication process: Under review

Manuscript of Study 2

Author	Author position	Scientific ideas %	Data generation %	Analysis & Interpretation %	Paper writing %
Yoana Omarchevska	First	85%	100%	90%	80%
Andreas Lachner	Second				
Juliane Richter	Third				
Katharina Scheiter	Fourth				

Title of paper: How do students' cognitive and motivational characteristics relate to experimentation skills and conceptual understanding during scientific inquiry?

Status in publication process: Published:

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*. https://doi.org/10.1080/10508406.2021.1966633

Manuscript of Study 3

Author	Author position	Scientific ideas %	Data generation %	Analysis & Interpretation %	Paper writing %
Yoana Omarchevska	First	85%	100%	90%	85%
Andreas Lachner	Second				
Juliane Richter	Third				
Katharina Scheiter	Fourth				

Title of paper: Do video modeling examples and metacognitive prompts improve self-regulated scientific inquiry?

Status in publication process: Published:

Omarchevska, Y., Lachner, A., Richter, J., & Scheiter, K. (2022). Do video modeling examples and metacognitive prompts improve self-regulated scientific inquiry?, *Educational Psychology Review*. https://doi.org/10.1007/s10648-021-09652-3

I. Introduction

Recent societal challenges that individuals face (e.g., COVID-19 pandemic, vaccines, energy supply, and climate change) require not only possessing conceptual knowledge about scientific topics but also understanding how scientific knowledge is created. Therefore, to actively participate in a technology- and science-based society, individuals need scientific reasoning and argumentation skills (OECD, 2016). Scientific reasoning and argumentation are essential for understanding scientific information presented in the media, for using scientific information during decision-making, for evaluating and justifying scientific arguments, and for participating in a democratic society (NASEM, 2016, 2021). The recent events of the COVID-19 pandemic further illustrate the importance of scientific reasoning and argumentation when evaluating scientific claims and evidence and when making decisions with personal and societal consequences (NASEM, 2021). For example, individual decision-making regarding vaccinating against COVID-19 or MMR (measles, mumps, and rubella) has individual but also societal consequences. To address these challenges, fostering the development of scientific reasoning became a fundamental aim of science education over the last decades (KMK, 2020, National Research Council [NRC], 1996; OECD, 2013). To achieve that, science education has moved towards instructional approaches which actively engage students in scientific reasoning activities (Kober, 2015).

The importance of developing scientific reasoning and argumentation skills (also referred to as scientific literacy or scientific competencies) has been continuously emphasized over the last decades (e.g., NRC, 1996, 2013) and by the latest science education recommendations (KMK, 2020). The newest recommendation from the National Academies of Sciences, Engineering, and Medicine (NASEM, 2021) in the United States, has positioned inquiry learning as the recommended instructional method to teach scientific reasoning and conceptual knowledge in science. During inquiry learning, students "learn science by doing science", they actively construct knowledge by conducting experiments, testing hypotheses, and engaging in discussions and argumentation (Next Generation Science Standards [NGSS], 2013). Thereby students can simultaneously gain conceptual knowledge and scientific reasoning skills (Hmelo-Silver et al., 2007). Having students actively participate in science aims to further increase their interest in science and inspire them to pursue future scientific careers. Furthermore, science curricula aim to foster students' abilities to conduct scientific investigations and reason in a scientific context (NRC, 2013). Likewise, the latest recommendation for science education in Germany, for example in biology, has also positioned scientific literacy as the main desirable outcome of secondary science education (KMK, 2020). Students are expected to gain conceptual understanding next to procedural competencies about different scientific methods. The educational standards in biology education comprise formulating questions, deriving hypotheses, planning and conducting investigations, evaluating findings, and drawing conclusions (KMK, 2020, p. 15). These competencies fall under the umbrella of scientific reasoning (Fischer et al., 2014) and are often taught following inquiry learning cycles (Pedaste et al., 2015).

Since educational policies have adopted inquiry learning as the preferred method of teaching science (NASEM, 2021; NRC, 1996), there are important factors that need to be considered when implementing such active learner-centered methods in science education. Active learning methods can lead to deeper conceptual knowledge and understanding and better learning outcomes by providing students with more freedom and control over their learning, but to do so, they rely on students' active participation and engagement (Freeman et al., 2014). Several factors might influence how students use the freedom they are provided with, how they approach scientific investigations, and which competencies they need to effectively learn from inquiry.

First, students' cognitive (e.g., conceptual knowledge) and motivational (e.g., interest, self-efficacy, self-concept) characteristics are particularly relevant for engagement and active learning (Huang, 2011; Simonsmeier et al., 2021). Students' motivation can impact their willingness and interest to engage with specific content like science (Renninger et al., 2018), whereas having conceptual knowledge can positively affect active learning (Hattie & Donoghue, 2016). Motivation can also influence students' perceptions of their ability to accomplish tasks associated with a specific domain (Renninger et al., 2018), for example conducting experiments in science. Furthermore, students' cognitive and motivational prerequisites are related to each other, for example, interest is related to engagement, knowledge gains, and self-concept (Tobias, 1994; Trautwein & Möller, 2016), suggesting that their combined effects need to be considered. Nevertheless, only the individual influence of these factors has been sparingly investigated in the context of inquiry learning.

Second, students' engagement is related to their ability to self-regulate the learning process, in terms of exhibiting the cognitive activation and behaviors associated with engagement (Wolters & Taylor, 2012). Inquiry learning requires students to plan investigations, monitor their comprehension, and reflect on the learning process, which are fundamental aspects of self-regulated learning (Schraw, 2010). For example, when performing multiple experiments, students need to monitor and evaluate whether the evidence they have collected is sufficient to answer the research question. Self-regulated learners can benefit more from the freedom that learner-centered learning approaches can offer compared to students who struggle with self-regulation (Azevedo et al., 2010; Greene et al., 2011). Because inquiry learning aims to foster scientific reasoning and argumentation, it is critical to consider the reasoning processes students engage in during inquiry. More importantly, because students need to regulate the inquiry process, the interaction between scientific reasoning and self-regulation processes requires further investigation and instructional support. Nevertheless, a detailed investigation of the conjoint use of self-regulation and scientific reasoning processes is yet missing.

Third, unguided inquiry learning methods (i.e., without providing instructional support) are associated with increased cognitive and metacognitive demands from students. Therefore, they have been continuously criticized because of the drawbacks of providing (too much) freedom (Kirschner et al., 2006; Mayer, 2004; Zhang et al., 2021). While freedom is necessary to achieve active learning, too much freedom might result in incorrect conceptual understanding. Being behaviorally engaged (for example, in discussions or experimenting) does not ensure cognitive engagement which is the kind of activity that leads to meaningful learning experiences (Mayer, 2004). Accordingly, several meta-analyses have concluded that, when appropriately guided, inquiry learning can be more effective than more traditional, expository teaching methods (Alfieri et al., 2011; Hmelo-Silver et al., 2007; Lazonder & Harmsen, 2016). The challenge of guided inquiry learning is finding the balance between the amount and type of guidance to provide (Mayer, 2004).

Previous research on guidance during inquiry learning has mainly focused on supporting students' scientific reasoning processes by structuring the inquiry process using prompts, scaffolds, heuristics (for a review, see Zacharia et al., 2015). However, fewer attempts were made to provide support for students' self-regulated learning (see, Manlove et al., 2009a, b). Moreover, if both self-regulated learning and scientific reasoning interplay, then the guidance provided should also foster their integration. Accordingly, in this dissertation I investigated the effectiveness of two types of guidance that aim to foster the self-regulation of scientific reasoning processes in an integrated manner.

To summarize, in this dissertation, I aim to advance our conceptual understanding of 1) the combined influence of students' individual prerequisites on experimentation and conceptual understanding, 2) the conjoint use of self-regulation and scientific reasoning processes in relation to argumentation, and 3) the effectiveness of guidance aimed at fostering the integration of self-regulation and scientific reasoning during inquiry learning. Three studies were conducted to address these three conceptual aims. Furthermore, an additional methodological contribution of this dissertation is going beyond focusing on learning outcomes and instead, providing an in-depth analysis of the conjoint, temporal use of students' learning processes using novel methods that model their interaction.

II. Scientific Reasoning and Argumentation

Scientific reasoning and argumentation skills are necessary for dealing with scientific information and understanding societal and scientific debates, for example, in the context of vaccines and climate change (NRC, 2013). Consequently, teaching scientific reasoning and argumentation skills have become fundamental aims of science education (OECD, 2013). In science education, students engage in tasks that require scientific reasoning to construct and evaluate scientific arguments (Osborne, 2010). The relationship between scientific reasoning and argumentation has different theoretical conceptualizations. Some theoretical frameworks have focused solely on scientific reasoning (e.g., Klahr & Dunbar, 1988), whereas others have focused only on scientific argumentation (Kuhn, 2010; Osborne, 2013). On the one hand, a more recent theoretical framework conceptualized scientific reasoning and argumentation together as a set of processes (Fischer et al., 2014). On the other hand, argumentation is directly related to scientific reasoning (Kuhn, 1993; Schwarz, 2009) and can be conceptualized as the outcome, or the product, of scientific reasoning tasks (Engelmann et al., 2016). For instance, when students are asked to write scientific explanations after conducting an inquiry task (e.g., performing investigations to learn about the principles of photosynthesis), the quality of the argumentation in their written answers can be seen as a sensitive indicator of the reasoning process. This dissertation adopts the latter conceptualization of argumentation as the product of scientific reasoning processes during inquiry tasks. In the following, scientific reasoning and argumentation are introduced separately and inquiry learning is introduced as a method to teach scientific reasoning and argumentation.

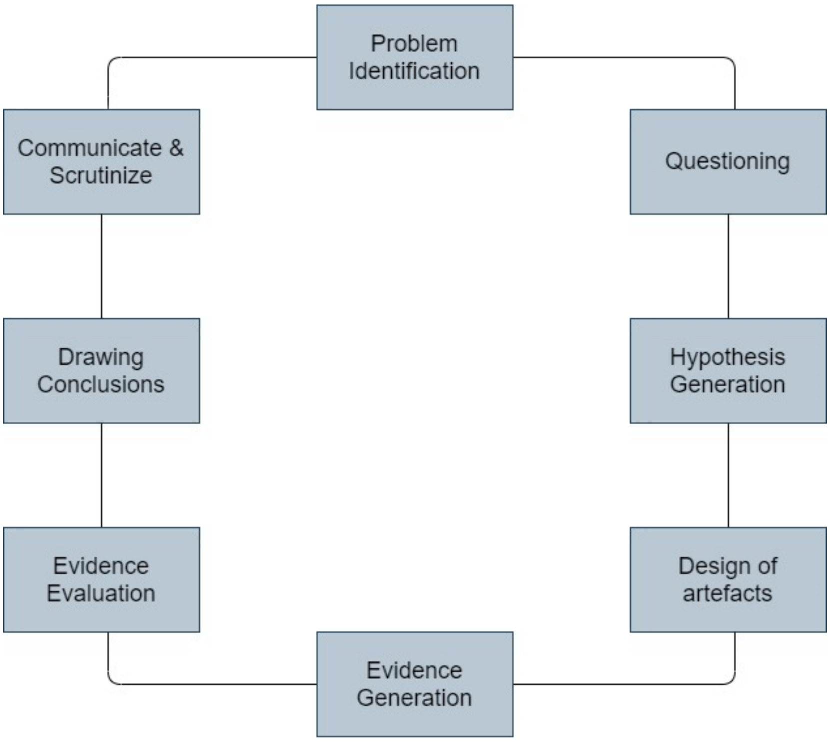
1. Scientific Reasoning

Scientific reasoning is defined as the skills and abilities to use and understand scientific methods and concepts and to evaluate and generate scientific knowledge (Engelmann et al., 2016; Fischer et al., 2014). Three broad groups of constructs have been identified in the context of scientific reasoning and argumentation – discovery process skills, argumentation skills, and understanding the nature of science (Engelmann et al., 2016). The commonality between these frameworks is the definition of scientific reasoning as an intentional problem-solving process that aims to coordinate evidence and theory (Klahr & Dunbar, 1988; Kuhn, 2010; C. Zimmerman, 2007).

Fischer et al. (2014) conceptualized scientific reasoning and argumentation as a set of cross-domain skills comprising eight epistemic activities (for a visual representation see Figure 1). The epistemic activities comprise different processes of scientific reasoning and argumentation – *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). Scientific reasoning is an iterative, non-linear process that typically starts with identifying a problem and building a problem representation (e.g., how do plants produce oxygen during photosynthesis?) and then distinguishing specific research questions (e.g., how does light intensity influence oxygen production?). Next, potential answers are derived in the form of testable hypotheses (e.g., if light intensity increases, oxygen production also increases). To test the hypotheses, evidence is generated following a deductive approach using theory-driven experiments, or an inductive approach using observations to make inferences (i.e., incrementally increasing light intensity in consecutive experiments).

Figure 1

A Visual Depiction of the Eight Epistemic Activities Identified by Fischer et al. (2014)



An important cognitive strategy for evidence generation is the control-of-variables strategy, which requires only manipulating the variable in question while keeping the remaining variables constant in an experimental trial (CVS; Chen & Klahr, 1999). Students' knowledge and use of the CVS is seen as a sensitive indicator of their ability to plan and conduct unconfounded experiments and establish causal relations between different variables. Scientific reasoning can also include designing and testing artefacts based on theoretical knowledge, for instance by creating a prototype in engineering (Fischer et al., 2014). After the evidence is generated, it is evaluated regarding the support it provides for the original claim (e.g., does the evidence support the hypothesis that increasing light intensity leads to increased oxygen production?). Last, conclusions are drawn by integrating different pieces of evidence and the findings are communicated and scrutinized by others during discussions. An integral process during communicating and scrutinizing is scientific argumentation. In the context of conducting experiments during inquiry learning, the most relevant epistemic activities are problem identification, hypothesis generation, evidence generation, evidence evaluation, and drawing conclusions. Therefore, this dissertation focused on these five epistemic activities.

In this context, *cross-domain* refers to the applicability of these skills in different domains and *skills* imply procedural knowledge about the process of scientific discovery (Hetmanek et al., 2018). Proposing a framework of epistemic activities provides a common vocabulary for discussing scientific reasoning and argumentation processes across domains and for facilitating the design of curricular interventions to foster their development (Hetmanek et al., 2018). Since this dissertation focuses on scientific reasoning processes, this framework (Fischer et al., 2014) was the basis for their fine-grained investigation and for the design of an instruction to support them. While the theoretical framework identified the different epistemic activities, no specific recommendations were made regarding the effectiveness of their sequential use (Csanadi et al., 2018). Consequently, no specific sequences of the epistemic activities are provided in the initial visual representation in Figure 1. To address this research gap, one aim of this dissertation is to investigate how different epistemic activities are used conjointly in relation to students' written argumentation. Namely, I investigated how the combined use of scientific reasoning processes relates to argumentation quality. For example, how does generating evidence and immediately evaluating it relates to argumentation quality at the end of the learning task?

2. Argumentation

Argumentation and critique are fundamental to the scientific method and have the purpose of presenting and justifying an argument to establish truth (Osborne, 2010). An argument typically consists of a claim, data, warrants, rebuttals, and qualifiers (Toulmin, 1958). Argumentation is tightly related to scientific reasoning and can have two general roles in science education. One line of research looks at argumentation as a dialogue (*arguing to learn*, Osborne, 2010) in classroom science, during which students can practice argumentation by engaging in argumentative collaborative discourse. During discourse, students can develop new understanding by presenting and justifying claims and by being challenged by their peers (Osborne, 2010). By engaging in a discussion with others, new ideas can be tested using rebuttals and counterarguments. Because a discussion requires collaboration with others, practicing argumentation as dialogue is more suitable for collaborative learning tasks.

Another line of research investigates students' written argumentation when providing scientific explanations in individual learning settings (Kuhn, 1993; Schwarz, 2009). For example, students engage in argumentation when drawing conclusions and answering a research question at the end of an inquiry learning task or when writing evidence-based explanations in science. As such, argumentation can be defined as the consequence, or the product, of scientific reasoning (Engelmann et al., 2016), and its quality can be used to assess students' scientific reasoning. One suitable model for evaluating argumentation quality in scientific explanations is the claim-evidence-reasoning (CER) framework proposed by McNeill et al. (2006). The CER framework breaks argumentation down into a claim (the conclusion answering the research question), evidence (the data provided to support the claim), and reasoning (the justification why the evidence is suitable to support the claim). The CER instructional framework was used as an intervention to support students in integrating theory and evidence into their written arguments and with evaluating principles against evidence. Therefore, next to conceptual knowledge, students can also learn epistemic knowledge, which broadly refers to how knowledge is created in science and what constitutes scientific evidence (Osborne, 2018). Because this dissertation focuses on scientific reasoning in individual learning tasks, this latter conceptualization of argumentation was adopted and the CER framework was used to evaluate students' argumentation quality.

3. Teaching Scientific Reasoning through Inquiry

Scientific reasoning and argumentation are complex skills and individuals of all ages often struggle with different aspects of them (de Jong & van Joolingen, 1998; Osborne, 2010).

Despite understanding how hypothetical beliefs differ from empirical evidence, children and adults have issues with articulating and evaluating evidence-based arguments (Fischer et al., 2014; Koslowski, 2012). Common difficulties with scientific reasoning include designing testable hypotheses and meaningful experiments (de Jong & van Joolingen, 1998) and relating theory and evidence (Smith & Wenk, 2006). Likewise, issues with argumentation involve not providing evidence for claims and justifications (Bell & Linn, 2000; Kuhn, 1991; McNeill, 2011). While engaging in scientific reasoning and argumentation is associated with difficulties, they can be improved with appropriate instruction and practice (Engelmann et al., 2016; Osborne et al., 2004). Scientific reasoning can be taught using different instructional methods, for example, using direct instruction or inquiry learning. What type of instruction is the most effective has been a topic of debate in research (Hmelo-Silver et al., 2007; Kirschner et al., 2006; Zhang et al., 2021) whereas educational policies encourage adopting more inquiry-based learning methods in science education (NRC, 2013). As a consequence, modern science education has shifted from using expository instructional methods to more active and learner-centered instructional methods like inquiry learning because they engage students in the epistemic activities associated with scientific reasoning (de Jong, 2019; Freeman et al., 2014).

Inquiry learning is an active learner-centered instructional method during which students learn science by performing experiments, making discoveries, and "acting like scientists" (Abd-El-Khalick et al., 2004; NRC, 2000). Thereby, students can simultaneously learn scientific reasoning skills and conceptual knowledge by conducting their own investigations (Furtak et al., 2012; Hmelo-Silver et al., 2007). Instead of passively learning about scientific facts, students are actively engaged in the scientific process, typically in five phases – orientation, conceptualization, investigation, conclusion, and discussion (Pedaste et al., 2015). Students generate hypotheses, plan and conduct experiments, collect and evaluate evidence, and draw conclusions (Lazonder & Harmsen, 2016; Pedaste et al., 2015). While experimentation has been traditionally implemented using hands-on experiments in physical science laboratories, virtual experimentation in online simulations is replacing or supplementing hands-on experimentation (Becker et al., 2020).

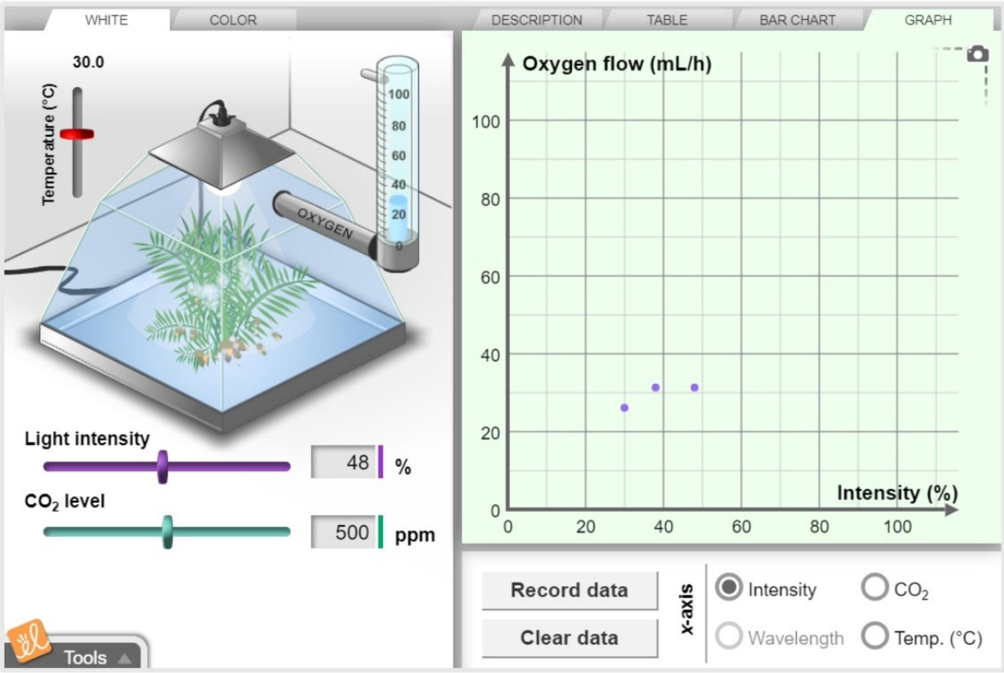
Learner-centered pedagogies like inquiry rely on students' active engagement and participation (Freeman et al., 2014). Students' individual prerequisites can influence their engagement and they are prominent factors for the effectiveness of these pedagogies (Chi & Wylie, 2014). Students' cognitive prerequisites, for example, the conceptual knowledge they possess at the beginning of a task, can influence the effectiveness of inquiry learning (Hattie &

Donoghue, 2016; van Riesen et al., 2018). Furthermore, students' motivational prerequisites, in terms of their interest and academic self-concept, can influence effort and achievement (Hidi, 1990; Huang, 2011). Although the importance of these factors has been established for learning in general (e.g., Marsh & Martin, 2011), their influence in the context of inquiry learning has received less attention in previous research. Moreover, students' cognitive and motivational prerequisites influence each other. For instance, interest is related to engagement, knowledge gains, and self-concept (Tobias, 1994; Trautwein & Möller, 2016), suggesting that their combined effects need to be considered (Kosel et al., 2021). Nevertheless, only the individual influence of these factors has been studied, and only sparingly, in inquiry learning research. Instead of focusing on the isolated influence of individual variables, person-oriented approaches have consistently demonstrated the importance of looking at the combined influence of students' individual prerequisites (Kosel et al., 2021; Seidel, 2006). The combined effects of students' cognitive and motivational prerequisites have not been investigated in the context of inquiry learning so far.

Modern inquiry learning is often embedded in computer-based learning environments like simulations using which students can perform virtual experiments (van Joolingen et al., 2007). Simulations are offered for different topics in the natural sciences (e.g., electric circuits in physics or photosynthesis in biology). Learning in such online environments has the benefit of providing a safe and controlled environment and allows for observing and investigating scientific phenomena which would otherwise be impossible (e.g., evolution). Using simulations, students can also explore simplified models and manipulate several variables and immediately see their effect (see Figure 2). For example, using a simulation as in Figure 2, students can investigate how different levels of light intensity affect oxygen flow during photosynthesis. Students can perform multiple experiments by incrementally increasing light intensity and note down the values of each experiment in a graph, table, or bar chart. In this way, computer-supported inquiry learning allows to study scientific reasoning processes in detail (e.g., evidence generation, evidence evaluation). In particular, further insight into the reasoning processes can be derived from analyzing logfile data of students' inquiry behavior or think aloud protocols (de Jong, 2006), for example, by modeling the sequences in which they manipulate variables.

Figure 2

A Simulation about Photosynthesis Created by Explore Learning



Note. Copyright (2021) by Explore Learning. Reprinted with permission.

Students' active participation and engagement in the knowledge construction process during inquiry learning can pose cognitive and metacognitive challenges (Azevedo, 2005; Scheiter & Gerjets, 2007). On one hand, difficulties can arise from students' lack of scientific reasoning skills, knowledge of inquiry processes, and a lack of conceptual knowledge. On the other hand, additional challenges can arise from a lack of metacognitive skills to regulate the learning process (de Jong & Njoo, 1992; Greene et al., 2008). Because of these challenges, some have argued that direct instruction should instead replace inquiry learning (Kirschner et al., 2006; Mayer, 2004; Zhang et al., 2021). At the same time, meta-analytic evidence consistently shows that inquiry learning can be more effective than expository teaching methods like direct instruction when appropriately guided using feedback, worked examples, or scaffolds (for meta-analyses, see Alfieri et al., 2011; Furtak et al., 2012; Lazonder & Harmsen, 2016). Therefore, research on inquiry learning has mainly focused on developing effective tools to guide the inquiry process (for a review, see Zacharia et al., 2015). Guidance can be provided in the form of heuristics, process constraints, direct presentation of information, scaffolds, or prompts (de Jong & Lazonder, 2014; Zacharia et al., 2015), which can be integrated into the online learning environments (van Joolingen et al., 2007).

Since both direct instruction and (guided) inquiry learning have their strengths, instead of comparing the two methods in a horse-race approach, in this dissertation, I investigate how both pedagogies can be best combined to support students' scientific reasoning and argumentation. Furthermore, because of the importance of self-regulation for active learning, alongside supporting students' inquiry processes, reasoning, and conceptual knowledge, it is important to also support their self-regulation (Zhang et al., 2004). Well-designed computer-supported inquiry learning environments should facilitate self-regulation and provide the necessary guidance for active learning (Shapiro, 2008; Zacharia et al., 2015). Moreover, because students need to self-regulate the inquiry process (e.g., monitor the suitability of evidence to test a hypothesis), integration of self-regulation and scientific reasoning is necessary. Nevertheless, studies on guidance during inquiry learning have mainly focused on either supporting the inquiry process (for a review, see Zacharia et al., 2015) or self-regulation (Manlove et al., 2007). First, this dissertation provides an in-depth investigation of the conjoint use of self-regulation and scientific reasoning processes in relation to learning outcomes during inquiry learning. Second, instruction that scaffolds the regulation of scientific reasoning activities is developed and tested on the learning outcome and process level.

III. Self-Regulated Learning

1. Definition of Self-Regulated Learning

Self-regulated learning is defined as the monitoring and regulation of cognitive, affective, metacognitive, and motivational processes before, during, and after learning (Azevedo et al., 2017; Schraw, 2010; B. J. Zimmerman, 2000). Several models of self-regulated learning have been proposed (e.g., Boekaerts, 1999; Pintrich, 2000; Winne & Hadwin, 1998; B. J. Zimmerman, 2013), which share the assumption that self-regulated learning is a dynamic, cyclical process consisting of three phases – preparatory, performance, and appraisal (Puustinen & Pulkkinen, 2001). Each phase comprises multiple processes, which can vary between different models; however, metacognitive events are an important aspect in almost all models (Engelmann & Bannert, 2021; Panadero, 2017). For instance, metacognition is a central aspect of self-regulated learning in the models by B. J. Zimmerman (2000) and Winne and colleagues (Winne & Hadwin, 1998; Winne & Perry, 2000).

Metacognition has two components – knowledge of cognition and regulation of cognition (Schraw & Moshman, 1995). Knowledge of cognition refers to understanding how one learns and involves declarative, procedural, and conditional knowledge about learning strategies and when to apply them (Schraw et al., 2006). Cognitive regulation[1] consists of four phases – planning, monitoring, control, and reflection (Pintrich, 2000). During the preparatory phase, students set goals for the learning task, and they plan their time and learning strategies (White & Frederiksen, 1998; Winne, 2001). Students' goals form the basis for planning and monitoring their learning progress. Monitoring is defined as students' awareness of their performance and understanding (Schunk, 2005). During the performance phase, students monitor their understanding and goal progression and adopt different cognitive strategies (i.e., control), if necessary (Nelson & Narens, 1990; Winne, 2001). Even though planning often occurs at the beginning of a learning task, students can adapt and reconsider their plans during the task based on the result of their monitoring activities (Hagemans et al., 2013). For example, if monitoring results in a lack of text comprehension, a strategy would be to reread or

[1] Pintrich (2000) defines regulation as the overarching term, whereas Nelson & Narens (1990) define monitoring and regulation as two distinct metacognitive processes.

summarize the text. The appraisal phase is concerned with evaluation and reflection on the learning process and the initial goals (B. J. Zimmerman, 2000).

Self-regulated learning is associated with greater academic achievement and learning gains (Azevedo et al., 2004; Land & Greene, 2000; Schunk & Greene, 2017). Self-regulated learning is particularly important in active learner-centered pedagogies like inquiry learning, which rely on students' active engagement and participation (Greene, 2018; Greene, Anderson, et al., 2018; Sinatra & Taasoobshirazi, 2018). Nevertheless, students often struggle with self-regulation (Schunk & Zimmerman, 2006), especially in difficult and unfamiliar tasks which impose high cognitive load (Seufert, 2018). For example, challenges are particularly likely to arise during learning in computer-mediated learning environments like simulations or hypermedia (Azevedo et al., 2010; Bannert, 2009; Bannert et al., 2015; Zheng, 2016). While self-regulation is associated with better learning outcomes in general (Schunk & Greene, 2017; Schunk & Zimmerman, 2006), successful learning in computer-based and learner-centered environments can largely depend on students' spontaneous or scaffolded self-regulation processes (Azevedo & Hadwin, 2005; Azevedo & Witherspoon, 2009; Bannert & Reimann, 2012; Dent & Koenka, 2016).

2. Self-Regulation during Inquiry Learning

To successfully learn from inquiry, students need to not only learn conceptual information but also to meaningfully navigate through the learning environment and follow scientific reasoning principles (van Joolingen & Zacharia, 2009). During inquiry learning students engage in complex scientific reasoning processes comprising multiple epistemic activities (van Joolingen & Zacharia, 2009) which require self-regulation (Manlove et al., 2009a). For example, students need to set goals and plan their investigations (White & Frederiksen, 1998), they also need to monitor whether they have collected sufficient evidence to draw conclusions, or whether the experiments they are conducting can test their hypotheses. Although monitoring and regulation are challenging during scientific reasoning tasks in general, self-regulation may become even more difficult during computer-based inquiry learning because of the learning environments' affordances.

Modern inquiry learning is often embedded in computer-supported learning environments like simulations using which students can perform virtual experiments (see Figure 2). Computer-based inquiry learning provides many affordances for learning complex topics in science and mathematics (Azevedo et al., 2017). For example, they provide students with multiple representations of information, allowing for a high level of control over their learning processes (Greene et al., 2011). While computer-supported learning environments provide several benefits for learning, such as performing multiple investigations in a short time and investigating various scientific phenomena (e.g., evolution, photosynthesis), they can also present additional challenges.

The benefits of conducting multiple experiments in a short time can be overridden when students engage in unsystematic experimental behavior. Namely, students might be tempted to try out several combinations of variables without a particular goal in mind ("gaming the system", Baker et al., 2008) and attempt to come to the correct solution, without following scientific reasoning principles (controlling variables and systematically collecting evidence). Because the costs of conducting a virtual experiment, in comparison to conducting a hands-on experiment, are lower, students might be tempted to simply act instead of think. In contrast, performing a hands-on experiment requires setting up the experimental environment which inherently requires some advance planning. Therefore, learning in computer-based inquiry environments can be associated with increased cognitive and metacognitive demands (de Jong, 2006). Self-regulatory processes like planning, monitoring, and employing different learning strategies are necessary to make use of these environments' affordances (Azevedo et al., 2010; Scheiter & Gerjets, 2007).

Since students' lack of self-regulatory skills poses difficulties during computer-supported inquiry learning (de Jong & Njoo, 1992; Greene et al., 2008), the learning environments should incorporate means to additionally support self-regulation (Shapiro, 2008; Zacharia et al., 2015). Consequently, considering self-regulation (and metacognition in particular) during inquiry learning has been emphasized in previous research (e.g., Andersen & Garcia-Mila, 2017; Kuhn, 2021; Manlove et al., 2007; Pedaste et al., 2012; White & Frederiksen, 1998; White et al., 2009). Providing students with metacognitive support, which facilitates self-regulation, can have a medium positive effect on students' performance, learning, and transfer (Azevedo et al., 2016; Bannert et al., 2015; Zheng, 2016). In particular, metacognitive expertise was shown to influence knowledge gains from inquiry learning (Chin & Brown, 2000; Kuhn et al., 2000; White & Frederiksen, 1998).

Accordingly, previous research has considered the influence of metacognition during inquiry either only theoretically (e.g., Andersen & Garcia-Mila, 2017), or has correlated self-regulation skills with the improvement of inquiry skills (e.g., Pedaste et al., 2012), or has directly supported self-regulation during inquiry (e.g., Manlove et al., 2007; White & Frederiksen, 1998, 2005). In fact, most research on self-regulation during inquiry learning has investigated the effectiveness of different self-regulation scaffolds, such as using a process coordinator (e.g., Manlove et al., 2007, 2009a, b) or concept maps (Hagemans et al., 2013). While these studies emphasize the importance of considering and supporting self-regulation (or metacognition) during scientific reasoning and inquiry learning, they also leave open several questions regarding the mechanisms through which self-regulation processes impact reasoning processes during inquiry learning. First, even though the effectiveness of self-regulation support during inquiry learning is investigated, students' self-regulation processes are often not considered and analyzed. Second, an analysis of how the scaffolds affect students' regulation of the inquiry process is also missing. Instead, findings from these studies are based on the isolated frequencies of self-regulatory processes which were correlated with measures of learning outcomes. However, to truly understand how self-regulation processes impact scientific reasoning processes and design scaffolds to support them, it is necessary to consider their interactions.

Previous research has demonstrated that integrating self-regulation scaffolds like a planning framework for the inquiry process (Njoo & de Jong, 1993; White et al., 1999) or prompts that support regulation (Veenman et al., 1994; Zhang et al., 2004) resulted in improved performance for students who received the scaffolds. Moreover, Manlove et al. (2006, 2007, 2009a, b) investigated the effects of regulative support scaffolds during individual and collaborative inquiry learning. In most studies (Manlove et al., 2006, 2007, 2009a) using regulation scaffolds was associated with higher learning gains but students did not make full use of the regulation scaffolds provided. In particular, students used the scaffolds for planning and goal setting, but not for monitoring. Likewise, three studies reported in Manlove et al. (2009b) found low instances of monitoring and that the use of the scaffolds was positively associated with lab scores but negatively associated with the quality of students' models. There are several potential explanations for the mixed findings reported above. On the one hand, Manlove et al. (2009b) suggest that the design of the scaffolds, requiring students to use an additional tool to engage in monitoring might impede students' tool use. They suggested that support measures which are integrated within the task might be more beneficial for monitoring activities. Nevertheless, to understand whether scaffolds integrating self-regulation and scientific reasoning are more effective, it is necessary to analyze their interaction during inquiry learning.

Conceptualizing self-regulation as a sequence of events (Bannert et al., 2014; Bernacki, 2018; Greene & Azevedo, 2010) and scientific reasoning as a set of processes (i.e., epistemic activities; Fischer et al., 2014), requires analyzing their temporal structures and interaction on a process level. Previous research has emphasized the importance of self-regulation for scientific reasoning and inquiry learning (Kuhn, 2021; Manlove et al., 2009b; White et al., 2009); nevertheless, an in-depth temporal investigation of how these processes co-occur during inquiry learning is yet missing. Methods that consider preceding and following events and can find patterns of events are needed (Winne, 2010). To comprehensively study how these constructs are related and design scaffolds to support them, it is fundamental to consider their co-occurrences and sequential relationships. In this dissertation, I address the existing research gap by analyzing the co-occurrences and sequential relationships between scientific reasoning and self-regulation processes and their impact on argumentation quality using two novel and advanced modeling methods.

IV. Temporal Modeling of Learning Processes

Process analyses, which model the temporal interactions between constructs, can provide a unique perspective and shed more light on the mechanisms through which self-regulation interacts with scientific reasoning and impacts argumentation quality. Implementing inquiry in computer-supported learning environments provides an opportunity to study both scientific reasoning and self-regulation processes by tracing students' actions using logfiles, think aloud protocols, and video recordings. These data sources incorporate temporal information about the timing and the duration of learning activities. Students' epistemic activities can be inferred from screen recordings (e.g., manipulating variables for investigations, use of the control-of-variables strategy), logfiles (information about clicks inside a simulation) or think aloud protocols (e.g., hypothesis generation, evidence evaluation). Likewise, self-regulation processes can be coded from logfiles or think aloud protocols which contain additional information about students' cognition and metacognition (Greene et al., 2018, 2021). Therefore, computer-supported (inquiry) learning environments can provide extensive, rich data about students' self-regulation and scientific reasoning processes (Azevedo et al., 2017; Winne, 2010). Moreover, because they contain information about the timing and sequence of different processes, they provide unique opportunities to study how scientific reasoning and self-regulation processes co-occur during learning. Regardless of their potential for advancing theoretical models, certain theoretical and methodological difficulties need to be considered – for example, different data sources need to be temporally aligned, novel and advanced methods are needed to analyze them, and accurate inferences about students' cognition and metacognition need to be made (Azevedo et al., 2017). In the next sections, I discuss the advantages of considering temporality in learning and suggest two process analyses that can help advance the field and tackle the challenges outlined by Azevedo et al. (2017).

Despite the rich affordances that process data can provide, their analyses are often limited to focusing on their frequencies, also known as a "coding-and-counting strategy" (Rack et al., 2018). During coding-and-counting, learning processes are coded, and their aggregated frequencies are used for group comparisons or as predictors of learning outcomes. While the aggregated frequencies of processes can provide information about their distribution, there are two main drawbacks associated with this approach (Csanadi et al., 2018; Reimann, 2009). First, when using aggregated data, it is not possible to account for the temporal development of activities (Reimann, 2009). Second, when data are aggregated, contextual information about preceding and following events of an activity and the specific patterns they exhibit over time is omitted and only their isolated occurrence can be quantified and compared (Reimann, 2009; Winne, 2010). In the context of investigating the self-regulation of scientific reasoning processes, one could either only compare the isolated frequencies of monitoring and evidence generation, for example, or one could also consider the events that triggered monitoring (or evidence generation) and the actions the learner performed after and therefore, examine *patterns* of events.

Consequently, a range of advanced statistical methods is being used in educational research (e.g., learning analytics, educational data mining). Recent research has shown the potential of process analyses to study self-regulated learning and recommends investigating the co-occurrences and temporal patterns of learning processes (Engelmann & Bannert, 2021; Lim et al., 2021). For instance, higher frequencies of metacognitive events are a necessary but not a sufficient precondition for initiating self-regulation (Azevedo & Hadwin, 2005; Bannert & Reimann, 2012; Bannert et al., 2015; Engelmann & Bannert, 2021). Using process data from learning activities, the temporal and adaptive features of self-regulated learning can be modeled during inquiry learning (Bernacki, 2018; Zhou & Winne, 2012). As a result, we can trace and model the temporal development of learning processes and shift the research focus from the traditional variable-based approach to an event-based approach (Molenaar, 2014; Reimann, 2009). While process analyses were mainly conducted in the field of self-regulated learning (e.g., Lim et al., 2021), defining scientific reasoning as a set of interrelated epistemic processes also affords a process-oriented investigation. In the following, two prominent methods for analyzing the co-occurrences and sequential relationships between scientific reasoning and self-regulation process are introduced – epistemic network analysis (ENA; Shaffer, 2017) and process mining (van der Aalst, 2016).

1. Epistemic Network Analysis

Epistemic network analysis (ENA) is a novel method that analyzes the connections among different elements by quantifying their co-occurrences and thereby accounts for their temporal relationships (Shaffer, 2017). The co-occurrence between two elements indicates that they are connected to each other in the same recent temporal context (Shaffer, 2018). ENA is based on the assumption that "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).

To analyze the structures of connections between different elements, ENA creates dynamic network models which quantify and visualize their co-occurrence (Shaffer et al., 2016). The network models illustrate the structure and the strength among elements making ENA suitable for answering a range of questions regarding patterns of associations between different elements (Shaffer et al., 2016). ENA quantifies the strength of association between different elements (i.e., individual codes) by modeling their frequency of co-occurrence (Shaffer, 2018). Different patterns which are observed at the same predefined timeframe can be identified, indicating that these processes are related to each other (Csanadi et al., 2018). Moreover, using ENA the strength of these co-occurrences can be compared in different groups. Thus, the temporal relationships between self-regulation and scientific reasoning can be compared in relation to argumentation quality or in different experimental conditions. These affordances make ENA particularly suitable for modeling how self-regulation and scientific reasoning processes interact during inquiry learning. While scientific reasoning and self-regulation have been scarcely and separately investigated using ENA, an analysis of their co-occurrences on a process level is missing.

In educational research, ENA has been previously applied in the context of diagnostic assessment in medical and teacher education (Bauer et al., 2020) or engineering education (Arastoopour et al., 2016). ENA was also used to compare individual and collaborative problem-solving with respect to epistemic activities (Csanadi et al., 2018). Csanadi et al. (2018) demonstrated the advantages of using ENA to uncover temporal relationships between epistemic activities in comparison to the traditional method of comparing the frequencies of codes. For example, their findings showed that dyads, compared to individuals, were more focused on evidence during their discussion, indicated by connections between evidence evaluation and all other epistemic activities. ENA has not been applied in the context of self-regulated learning, apart from a study by Wu et al. (2020) who compared patterns of metacognitive phenomena between high- and low-scoring students and between students from different disciplines (science or humanities). Their findings showed that high-scoring students had stronger connections between goals and actions, whereas low-scoring students focused on metacognitive knowledge and context. These results were obtained from self-reflection reports which are more subjective and may not fully capture metacognition (Wu et al., 2020). Therefore, the authors suggest combining different data sources of metacognition (video data, think aloud protocols, self-reports) for a more comprehensive analysis of students' metacognitive processes.

ENA is a particularly suitable method for relating different sources of process data. For example, combining think aloud data with other video or screen recordings can support making better inferences about students' cognitive and metacognitive processes (cf. Azevedo et al., 2017). To further support the inferences from these analyses, the specific connections identified by ENA can be easily validated by going back to the original data and therefore "closing the interpretative loop" (Shaffer, 2017). In this way, quantitative relationships can be supported qualitatively, and the reliability of the identified networks can be assessed. Therefore, ENA is one novel and advanced analytical approach that can help address the challenges outlined by Azevedo et al. (2017).

2. Process Mining

Process mining is an educational data mining method that uses logfiles to create process models describing the underlying relationships between different events (van der Aalst, 2016). These process models can provide more insight into how learning processes unfold by considering the temporal aspect of learning and focusing on the patterns of cognitive and metacognitive processes (Engelmann & Bannert, 2021). To achieve that, the timing of learning events is used to extract sequences and visualize the preceding and following events in the form of process models. Next to visualizing the relationships between different events, process mining also quantifies their relationships using metrics like frequency, dependency, or timing (Saint et al., 2021). Different process mining algorithms (e.g., heuristics miner, fuzzy miner) can be used to identify the relationships between processes based on their sequences and associations. Processes can be obtained from digital traces (logfiles, trace data) or coded data like video recordings or think aloud protocols. Importantly, timestamps of the processes are necessary to derive the sequences between different processes. Then, the transitions between the processes are used to create process models which quantify and visualize the relationships between them (Saint et al., 2021).

Process mining's affordances for modeling the temporal relationships between learning processes resulted in its increased use in current self-regulated learning research (Engelmann & Bannert, 2021; Lim et al., 2021; Sonnenberg & Bannert, 2019). The conceptualization of self-regulated learning as a cyclical, temporal process requires methods that can account for temporality and extend beyond frequency measures (Molenaar, 2014; Reimann, 2009). For example, process mining was used to model the temporal structure of self-regulated learning processes obtained from coded from think aloud protocols (Bannert et al., 2014; Lim et al., 2021; Sonnenberg & Bannert, 2015, 2019). Engelmann & Bannert (2021) demonstrated that considering only the frequencies of metacognitive and cognitive events is less meaningful than a sequential analysis on the process level. While process mining is being more consistently used in self-regulation research, it has not yet been applied in the context of scientific reasoning or the regulation of scientific reasoning processes.

Process mining can be used to align different data channels, for instance, video recordings and think aloud protocols when these channels have the same time frames (e.g., hh:mm:ss). Moreover, combining data sources of students' cognitive and metacognitive processes (e.g., obtained from think aloud protocols) with behavioral sources of their problem solving (e.g., obtained from video recordings) allows for more accurate inferences of these processes, compared with relying on a single data source (cf. Azevedo et al., 2017). Therefore, process mining is particularly suitable for providing unique insights into the sequential relationships between self-regulation and scientific reasoning processes.

To conclude, ENA and process mining each provide a unique insight into the co-occurrences and sequential relationships between scientific reasoning and self-regulation processes. ENA is particularly suitable for visualizing and comparing the strength of the co-occurrence between scientific reasoning and self-regulation processes in different groups of students. For example, using ENA it would be possible to identify whether planning and evidence generation co-occurred more frequently in students with higher argumentation quality. However, ENA does not provide information about the sequence of the two processes, namely whether planning was performed before or after evidence generation. The specific sequence can, however, be identified using process mining, indicating that the two methods can complement each other. Nevertheless, the two methods have been only used separately in prior research. To address this research gap, in this dissertation, I first used ENA alone and later combined it with process mining for an in-depth analysis of the relationships between scientific reasoning and self-regulation processes.

V. Objectives and Expected Outcomes

This dissertation has three overarching conceptual aims. The first aim is to investigate the combined influence of students' cognitive and motivational characteristics on students' inquiry behavior and conceptual understanding. Study 1 (reported in section VI) related students' cognitive and motivational characteristics to their experimentation skills, conceptual understanding, and argumentation during inquiry learning. The second aim is to provide an in-depth investigation of the interplay between self-regulation and scientific reasoning on a process level in relation to argumentation quality (Study 2, reported in section VII). The third aim of this dissertation (Study 3, reported in section VIII) is to design an intervention based on the findings from Study 2 that scaffolds scientific reasoning and self-regulation in an integrated manner and to test its effectiveness on the process level (i.e., how students engaged in these processes after instruction) and the learning outcome level (i.e., students' hypothesis and argumentation quality).

An additional methodological contribution of this dissertation relates to addressing the three conceptual aims using a range of novel and advanced methods. Each of the three studies reported in this dissertation employed a novel methodological approach to provide a fine-grained analysis of students' learning processes. First, in contrast to the classic variable-oriented perspective often used in educational research, a person-oriented approach was applied to understand the combined influence of cognitive and motivational variables on students' experimentation skills and conceptual understanding. Second, epistemic network analysis was used to model the temporal co-occurrence of self-regulation and scientific reasoning processes in relation to argumentation quality. Third, epistemic network analysis was combined with process mining to comprehensively investigate specific sequences and co-occurrences between scientific reasoning and self-regulation processes. Process data were used in all three studies to gain an in-depth understanding of students' scientific reasoning and self-regulation in relation to learning outcomes during inquiry learning. Therefore, each study goes beyond looking only at learning outcomes and instead focuses on providing an in-depth process-oriented perspective on the interplay between scientific reasoning, self-regulation, and argumentation. The following overarching research questions guided this dissertation:

RQ1) How do students' cognitive and motivational characteristics influence experimentation skills and conceptual understanding during inquiry learning (Study 1)?

RQ2) What is the role of self-regulation processes for scientific reasoning and argumentation during inquiry learning (Study 2)?

RQ3) How can self-regulation and scientific reasoning during inquiry learning be scaffolded (Study 3)?

VI. Study 1: How do Students' Cognitive and Motivational Characteristics Relate to Experimentation Skills and Conceptual Understanding during Scientific Inquiry?

1. Abstract

This study used a person-centered approach to investigate how cognitive and motivational characteristics relate to experimentation skills and conceptual understanding during guided inquiry learning. All students (N = 110, M_{age} = 12.07) attended a lesson about photosynthesis and worked on a structured inquiry task individually. Experimentation skills were measured by the proportion of controlled experiments. Conceptual understanding comprised students' written argumentation and posttest scores. We identified three distinct clusters, based on students' prior knowledge, self-concept, and interest – Underestimating, Struggling, and Strong. Although Struggling students conducted a significantly higher proportion of controlled experiments than Strong students, Underestimating and Strong students outperformed them in the conceptual knowledge posttest. Using experimentation skills to explain differences in conceptual understanding is not suitable in structured tasks where students have no freedom to experiment. We found no differences in argumentation quality. Students' interest and self-concept need to be considered when explaining the effectiveness of inquiry learning.

Publication process: Under review: Learning and Instruction

2. Theoretical Background

Inquiry learning is often used in science education because it actively engages students in hypothesis generation, experimental design, and argumentation when learning about scientific phenomena (Furtak et al., 2012). To be successful, students need experimentation skills, which can influence their conceptual understanding (Bryant et al., 2015). For example, to design unconfounded experiments, students should only manipulate one variable at a time and keep the remaining variables constant (control of variables strategy (CVS); Chen & Klahr, 1999). With appropriate guidance, learning from inquiry can be more effective than direct instruction (Lazonder & Harmsen, 2016). Nevertheless, even when guidance is provided, students can vary in conceptual understanding (van Riesen et al., 2019), suggesting that inter-individual differences may be a boundary condition for the effectiveness of (guided) inquiry learning (de Jong, 2019).

As active-learner pedagogies rely on students' engagement and activity, it is important to consider how motivational and cognitive factors influence students' experimentation skills and conceptual understanding. Pedagogical approaches like inquiry learning, that require more active monitoring and control by learners, are likely more affected by features related to the ability to control, namely cognitive and motivational prerequisites (Scheiter & Gerjets, 2007). Previous research has studied the effects of cognitive and motivational variables from a variable-centered perspective (Jansen et al., 2015; Mulder et al., 2009; van Riesen et al., 2019). In contrast, in the present study we applied a person-centered approach to first identify subgroups of students based on their configurations of multiple characteristics (e.g., knowledgeable but unmotivated) and second to relate the subgroups resulting from this approach to learning processes and outcomes (Seidel, 2006).

2.1 Inquiry Learning

Modern science education emphasizes the need to involve students in the knowledge construction process by implementing more active learner-centered instruction (de Jong, 2019). Active learning can result in deeper understanding and better learning outcomes compared to more passive instructional methods (Freeman et al., 2014). Inquiry learning actively engages students in the scientific process and relies on their participation (de Jong & van Joolingen, 1998). During inquiry learning, instead of receiving direct instruction about scientific facts, students develop hypotheses, conduct experiments, and evaluate findings to derive conclusions about the underlying scientific principle in question (Lazonder & Harmsen, 2016; Pedaste et al., 2015). In this way students can gain conceptual, procedural, and epistemic knowledge (Furtak et al., 2012).

Inquiry learning is typically divided into inquiry phases which form an inquiry cycle (Pedaste et al., 2015; White & Frederiksen, 1998). Pedaste et al. (2015) synthesized different inquiry phases and cycles into an inquiry framework consisting of five distinct phases – orientation, conceptualization, investigation, conclusion, and discussion. During the orientation phase, the problem is identified, and the important variables are specified. The conceptualization phase involves two sub-phases, questioning and hypothesis generation, during which the relationship between different variables is stated. The product is a research question and/or specific hypotheses about the relationship between different variables. The investigation phase consists of three sub-phases: exploration, experimentation, and data interpretation. In the exploration and experimentation phases, the experiment is designed, and data are collected. Then, data are interpreted, and new knowledge is synthesized. During the conclusion phase, the findings are interpreted and a final conclusion made. The discussion phase involves reflection about the inquiry process after each phase or communicating the findings to a broader audience.

Due to its complexity students often struggle with different aspects of the inquiry process, like designing meaningful experiments and interpreting data (de Jong & van Joolingen, 1998). Inquiry learning is particularly effective when additional support is provided (Furtak et al., 2012; Lazonder & Harmsen, 2016). Nevertheless, the effectiveness of guided inquiry learning may depend on students' prior knowledge (van Riesen et al., 2019), experimentation skills (Bumbacher et al., 2018), and interest (Renninger et al., 2014). To be able to conduct and interpret scientific experiments, a fundamental experimentation skill is the control of variables strategy (CVS; Chen & Klahr, 1999). The CVS entails that, to produce unconfounded results, only the independent variable should be manipulated in an experimental trial and the remaining variables should be kept constant (Chen & Klahr, 1999). An important predictor of the CVS is students' science content knowledge in secondary education (Schwichow et al., 2020). Despite its central role for successful inquiry, students and adults often struggle with the CVS (Zimmerman & Croker, 2013) by manipulating irrelevant variables or multiple variables at once (de Jong, 2006). Providing students with guidance was effective for learning the CVS, for example, using video modeling examples (Kant et al., 2017) or task structuring (Lazonder & Kamp, 2012).

When assessing experimentation skills, it is more meaningful to measure hands-on behavior than using multiple-choice or open-ended questions (Schwichow et al, 2016). Stender et al. (2018) showed that only experimentation skills, measured as controlled experiments and hypothesis evaluation, were directly related to knowledge gains, whereas declarative knowledge about the CVS had only an indirect effect via experimental skills, which implies that even though students understand the CVS in theory, they do not necessarily apply it in practice. Therefore, we coded screen captures of students' inquiry behavior as a hands-on measure of the CVS (Kant et al., 2017).

2.2 Conceptual Understanding

Students' learning outcomes at the end of inquiry are typically measured using tests on conceptual understanding (e.g., Husnaini & Chen, 2019). Because inquiry aims to generate and justify knowledge claims, argumentation is another important indicator of conceptual understanding (e.g., Jimenez-Aleixandre et al., 2000). Argumentation aims to establish the relationship between claims and evidence, positioning itself in the center of scientific inquiry (Duschl & Osborne, 2002). For example, students engage in argumentation when they interpret evidence, but they often use inappropriate or irrelevant data in their arguments (Sandoval & Millwood, 2005). Students also need to provide reasoning in their argumentation; however, they often provide vague or purely descriptive explanations without specifying the underlying causal mechanism (McNeill et al., 2006). Prior content knowledge is a strong indicator of producing high-quality arguments as students often build upon their prior knowledge and experiences when engaging in argumentation (von Aufschnaiter et al., 2008). We used argumentation quality, alongside test scores, as a measure of conceptual understanding. We adapted the claim-evidence-reasoning framework (McNeill et al., 2006) to our context. The claim is the answer to the research question, the evidence is the provided data to support the claim, and the reasoning justifies using the evidence to support the claim.

2.3 Individual Characteristics and Inquiry Learning

Cognitive and motivational characteristics are highly predictive and relevant for learning, engagement, and academic achievement (Huang, 2011; Simonsmeier et al., 2021). Engagement is particularly important in active pedagogies and is associated with better learning outcomes (Chi & Wylie, 2014). For example, interest, self-concept and prior knowledge can affect students' engagement and effort (Hidi, 1990; Rodrigues, 2007; Schnitzler et al., 2021). In the following, we discuss the individual role of these factors for inquiry learning.

2.3.1 Prior Knowledge. Active forms of learning, like inquiry learning, are more effective for obtaining deep understanding when learners already possess a certain level of prior knowledge (Hattie & Donoghue, 2016). However, to benefit learning, prior knowledge needs to be activated, relevant, and congruent (Brod, 2021). To be able to generate hypotheses about the relationships between variables and design experiments to test them, students need at least some prior content knowledge. Students employ different investigative strategies depending on their levels of prior knowledge (Lazonder et al., 2008). High prior knowledge is associated with generating and testing specific hypotheses, better performance (Lazonder et al., 2008), and more goal-directed behavior (Hmelo et al., 2000). Furthermore, high-prior knowledge students use more sophisticated strategies during inquiry learning (Schauble et al., 1991), for example, the CVS (Kanari & Millar, 2004). Low-prior knowledge students have difficulties relating different variables or they simply guess their relationships (Mulder et al., 2009). They also tend to design simpler experiments, manipulate irrelevant variables, neglect control variables (de Jong, 2006), and violate the CVS more frequently than students with high prior knowledge (Bumbacher et al., 2018).

2.3.2 Academic Self-Concept. Academic self-concept refers to one's perceived ability in an academic domain (Marsh & Martin, 2011) and has a positive reciprocal relationship with academic effort (Trautwein et al., 2006), achievement (Huang, 2011), choice of academic courses and interest (Trautwein & Möller, 2016). Students who perceive themselves as competent in an academic domain are likely to be more successful than their peers with similar abilities (Marsh & Martin, 2011). Consequently, academic self-concept is a prominent determinant of achievement and fostering academic self-concept is a desirable educational outcome (OECD, 2003).

Academic self-concept is domain-specific, because while students might feel competent in one subject (e.g., English), they can have a low self-concept in another subject (e.g., science). Science academic self-concept is a multidimensional, subject-specific construct (Jansen et al., 2014). Subject-specific grades are strong predictors of academic self-concept in the same subject, but not in other subjects (Jansen et al., 2014). Reciprocally, academic self-concept is a significant predictor of science achievement when considered separately from self-efficacy (Jansen et al., 2015).

Engaging in inquiry activities has overall positive effects on academic self-concept, suggesting that it is important to foster positive mastery experiences during inquiry to avoid feelings of failure which negatively impact students' self-concept (Jansen et al., 2015). While previous meta-analyses have shown a reciprocal relationship between achievement and academic self-concept (Huang, 2011; Marsh & Martin, 2011), an analysis of how academic self-concept influences students' experimentation and conceptual understanding during inquiry is yet missing. Outside of the context of inquiry, higher academic self-concept was related to higher engagement and higher achievement in the end of the year (Schnitzler et al., 2021).

2.3.3 Interest. Interest is conceptualized as the psychological state of engaging, and the motivation to re-engage, with specific content (Hidi & Renninger, 2006). Interest is a multidimensional construct comprising affect, knowledge, and value regarding a performed activity (Hidi & Renninger, 2006). Having interest in a specific content (e.g., science) is associated with deeper understanding, greater effort, and perseverance despite facing difficulties (Hidi, 1990). Moreover, interest has a positive effect on students' attention, goal setting, and the development and use of learning strategies (Renninger & Hidi, 2011). In science learning, interest has reciprocal links with the environmental context (science classes) and students' behavior (Glynn et al., 2015).

Motivation can guide and sustain science learning behaviors (Glynn et al., 2015). Students' situational interest can be fostered by engaging them in novel, challenging, and social activities, like inquiry (Renninger et al., 2014). For example, an inquiry-based intervention positively affected students' science learning (Renninger et al., 2014). Importantly, the intervention had positive effects for students with different levels of interest and suggested that students with varying levels of interest might need different types of support. While the benefits of inquiry on generating and maintaining students' interest in science have been investigated in prior research, the evidence regarding the influence of pre-existing interest in science on students' experimentation skills and conceptual understanding is scarce.

2.4 Combined Effects of Student Characteristics

The studies discussed above relate the individual effects of cognitive and motivational characteristics on students' learning outcomes, by applying a variable-centered approach (Howard & Hoffman, 2018). This approach, however, measures the influence of student prerequisites in isolation, and ignores that, for instance, cognitive and motivational characteristics are consequently interwoven (Tobias, 1994). As such, it is important to investigate their combined effects (Kosel et al., 2021). For example, interest is related to engagement and knowledge acquisition (Tobias, 1994) and self-concept is related to interest (Trautwein & Möller, 2016). Recent studies, therefore, adopted a person-centered approach, which particularly takes the dependencies of student prerequisites into account, by identifying unique subgroups of students based on their cognitive and motivational characteristics (Kosel et al., 2021; Schnitzler et al., 2021; Seidel, 2006). The person-centered approach uses methods like cluster analysis and mixture modeling to identify unique subgroups (high within-group similarity and high between-group variability) of participants based on certain grouping variables (Howard & Hoffman, 2018).

The value of a person-centered analysis lays in the identification of heterogeneous subgroups as well as homogeneous subgroups (Seidel, 2006). Kosel et al. (2021) classified more than half of the students in their sample into heterogeneous groups, characterized by incoherent combinations of cognitive and motivational characteristics (e.g., knowledgeable but unmotivated). Such analysis can provide educational implications for the needs of diverse students (Hayenga & Corpus, 2010). A person-centered analysis has not been yet applied in relation to learning outcomes and processes during inquiry. Therefore, by using a person-centered approach, we investigated how cognitive and motivational variables relate to secondary students' experimentation skills and learning outcomes.

3. The Present Study

We present a field study which investigated how students' individual prerequisites relate to experimentation skills and conceptual understanding during computer-supported inquiry learning. The study was conducted in a technology-enhanced laboratory mimicking a classroom (TüDiLaB) in which data can be collected with whole classes. The laboratory is equipped with laptops, tablets, and internet and provides a controlled setting for conducting technology-enhanced lessons involving inquiry learning. Students' individual prerequisites (self-concept, interest, and prior knowledge) were assessed prior to attending a biology lesson in the lab. During the lesson, all students solved a computer-supported inquiry task on the topic of photosynthesis. At the end of the inquiry task, students wrote an answer to the research question they investigated, from which we coded argumentation quality. Then, they solved the conceptual knowledge posttest. We preregistered the following research questions and analyses (https://aspredicted.org/blind.php?x=/YX5_T7J). Because we did not preregister specific hypotheses, we only state our expectations:

RQ1) What types of clusters can we identify based on students' individual prerequisites?

Because mixture modeling (e.g., latent profile analysis) requires larger samples, we used cluster analysis to distinguish students based on their levels of prior knowledge, academic self-concept, and interest in biology. We expected to identify maximum 4 clusters with different combinations of these variables. For example, groups with homogeneous scores on all variables, and groups with heterogeneous scores on the cognitive and motivational variables (Kosel et al., 2021).

RQ2) Do the clusters differ in experimentation skills?

We measured experimentation skills as the proportion of controlled experiments (CVS) coded from screen captures of students' inquiry behavior. In line with Schwichow et al. (2020), we expected that high-prior knowledge students would perform more controlled experiments.

RQ3) Do the clusters differ in conceptual understanding?

We compared the clusters in conceptual understanding, measured by students' posttest scores and the argumentation quality in their written answers. We expected that students with high prior knowledge, interest, and self-concept to have better conceptual understanding.

RQ4) Do experimentation skills mediate the relationship between individual prerequisites and conceptual understanding?

Previous research (Bumbacher et al., 2018) has shown that experimentation strategies are strongly and positively correlated with factors of conceptual understanding. Therefore, we investigated whether experimentation skills mediate the relationship between individual characteristics and conceptual understanding.

4. Method

4.1 Participants and Design

Participants were 110 seventh grade students from six[2] classes in Southern Germany. Students were 12.07 years old on average (*SD* = 0.56) and 57.3% of them were male. Students were native German speakers (71.8%), bilingual (9.1%), or had another mother language (15.5%), and one participant did not state their mother language. On average, participants' last grade in biology was 2 (*SD* = 0.70), ranging from 1 (*very good*) to 6 (*failed*).

The study was a part of a larger research project which also addressed the effects of an additional experimental manipulation that was, however, administered after the lesson and the inquiry task (Jacob et al., 2022). Students in all classes and independent of this experimental manipulation received the same instruction and performed the same inquiry task, which is the focus of the present study. Data relating to this task have not been analyzed previously. The research project was approved by the local government and informed consent for participation was obtained from all students and their parents.

4.2 Measures

4.2.1 Academic Self-Concept. Academic self-concept in biology was measured using 6 items from PISA (OECD, 2016; Cronbach's α = .72), rated on a Likert scale ranging from 1 (*completely disagree*) to 4 (*completely agree*). An example item of the scale is "I quickly learn new material in biology".

4.2.2 Interest. Interest in biology was measured using five items from PISA (OECD, 2016; Cronbach's α = .88), rated on a Likert scale ranging from 1 (*completely disagree*) to 4 (*completely agree*). An example item of the scale is "I am interested in learning something new in biology."

4.2.3 Prior Knowledge. The test assessed prior knowledge about different principles of photosynthesis (e.g., "What is the function of chloroplasts?") using five multiple-choice items with one correct answer (maximum score = 5). Because the items assessed knowledge about various aspects of photosynthesis, unsurprisingly the internal consistency was low

[2] The original sample comprised 7 classes (Jacob et al., 2022). However, after collecting data from the first class, the instruction and inquiry task were shortened. Therefore, we analyzed the six remaining classes.

(Cronbach's α = .37; see Stadler et al., 2021, for the inadequacy of Cronbach's alpha as a quality indicator for knowledge tests).

4.2.4 Experimentation Skills. We coded experimentation skills from screen captures of participants' actions during the inquiry task. For each experiment conducted, we coded the use of CVS, correct variables, and incorrect variables. Experimentation skills were defined as the proportion of controlled experiments relative to the total number of experiments (Kant et al., 2017). We used the CVS as a measure of experimentation skills because it incorporates information about the manipulation of all variables in each distinct experiment, which is more informative than the isolated frequencies of correct or incorrect variable use (see Table 3 for descriptive statistics). To establish inter-rater reliability, two trained raters independently coded 20% of the data (21 videos, 1380 codes) and reached perfect agreement (κ = .99); one of the raters coded the remaining videos.

4.2.5 Argumentation Quality. We asked students to synthesize the previously collected evidence and provide a short explanation regarding the question "*How are the CO_2 concentration and the oxygen production related?*". Students' answers had 22 words on average (*SD* = 13.7). Argumentation quality was assessed by scoring by participants' answers using an adapted version of the coding scheme from McNeill et al. (2006), see Table 1. The scheme assessed the quality of the claim (0-1), the evidence (0-1), and the reasoning (0-1), adding to a maximum score of 3. Two raters independently scored 20% of the answers (*n* = 20) and achieved perfect inter-rater reliability (κ = .93).

4.2.6 Conceptual Knowledge Posttest. In the posttest, conceptual knowledge was measured using nine different items (Cronbach's α = .58). Each question had one correct answer (maximum score of 9). An example item is "*Which substance is the product of photosynthesis?*".

4.3 Lesson on Photosynthesis

All classes attended a scripted lesson about photosynthesis taught by the same teacher. Students first filled out a paper questionnaire assessing their demographic information, prior conceptual knowledge, academic self-concept, and interest. We co-designed the lesson with pre-service biology teachers. Following the inquiry learning cycle (Pedaste et al., 2015), the lesson had five phases. First, students watched an introductory video about photosynthesis to re-activate their knowledge (*orientation* phase). Second, students received direct instruction on the components of photosynthesis (*conceptualization* phase). The teacher introduced the *Photosynthesis* simulation (Figure 3) and explained how different environmental conditions can be simulated (e.g., light intensity can be varied with a light bulb). The teacher explained the inquiry task and answered questions.

Table 1

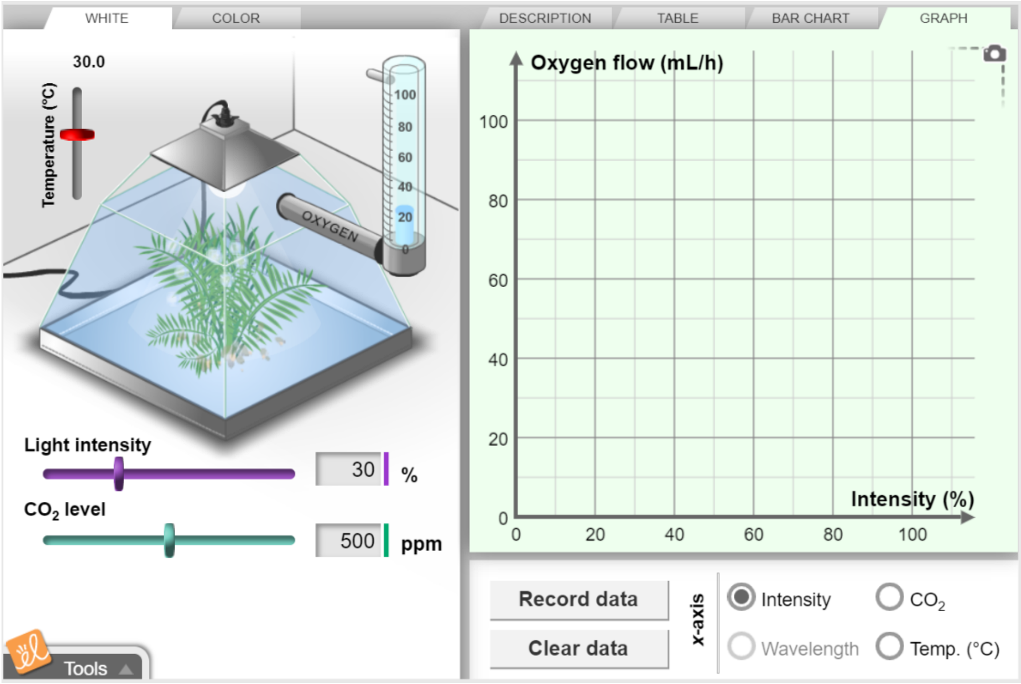
Coding Scheme for Assessing Argumentation Quality, Adapted from McNeill et al. (2006)

Component	*0*	*1*
Claim	No claim or an incorrect claim about the relationship between CO_2 oxygen production *"If you give the plant more oxygen, the plant returns more CO_2." (reverse relationship)*	Makes an accurate and complete claim. *"The more CO_2, the higher the oxygen formation rate."*
Evidence	No evidence or wrong evidence *"Goes in 5-degree steps"*	Provides concrete and correct evidence *"But above a certain CO_2 level (300ppm) the oxygen rate remains the same."*
Reasoning	Provides incorrect scientific principle or reasoning that does not link the claim and data. *"The more CO_2 is produced, the more photosynthesis the plant does."*	Mentions the correct scientific principle *"Oxygen is produced by the CO_2. Without CO_2 there would be no oxygen. The more CO_2 concentration, the higher the oxygen formation rate."*

Third, students used a structured worksheet and the *Photosynthesis* simulation to conduct experiments and collect data individually (*investigation* phase). The worksheet guided students' experimentation and use of the CVS by providing specific values for temperature, light intensity, and CO_2. Each class was split in six groups and each group was given different values for light intensity and temperature and the same six values for CO_2. In the end, there were six different combinations of light intensity and temperature values that the groups investigated for six CO_2 values. For example, one group had to keep the values 21° C and 20% light intensity constant and vary six CO_2 levels. The resulting values for oxygen flow were filled in a table and then transferred to a graph on the worksheet, which was identical to the graph in the simulation (see the right section of Figure 3).

Figure 3

The Photosynthesis Simulation Used in the Inquiry Task



Note. Copyright (2021) by Explore Learning. Reprinted with permission.

Last, during the *conclusion* and *discussion* phases, the teacher collected the answers from each group and noted the correct answers on the smartboard in a whole-class discussion. All students could see how different levels of light intensity, temperature, and CO_2 affect oxygen flow. Conclusions about the individual effects of light intensity and temperature were drawn. Students wrote a short interpretation of their findings about the relationship between CO_2 concentration and oxygen flow, which was not previously discussed with the class. Afterwards, all students wrote a mnemonic on their worksheets, explaining the relationship between the three factors. The duration of the lesson was about 45 minutes. Then, another experimental study followed (Jacob et al., 2022), which did not influence the outcomes addressed in this paper.

4.4 Data Analysis

To obtain the optimal number of clusters (RQ1), we performed a *k*-means cluster analysis in RStudio (RStudio Team, 2020). Since the input variables (self-concept, interest, and prior knowledge) for the cluster analysis were measured on different scales, they were first *z*-standardized. The optimal cluster solution was determined by inspecting the scree plot, which illustrates the within-cluster sum of squares as a function of the number of clusters (1-15). The largest decrease in within-cluster sum of squares indicates the optimal number of clusters (López-Rubio et al., 2018). Next, we used one-way ANOVAs and post-hoc comparisons with Bonferroni corrections to compare the clusters in experimentation skills (RQ2) and conceptual understanding (RQ3). In case of assumption violations, we used Welch-ANOVA and Games-Howell corrections. We additionally computed learning gains, measured by the difference in average correctly solved questions between posttest and pretest. To determine whether experimentation skills mediate (RQ4) the relationship between individual characteristics (self-concept, interest, and prior knowledge) and conceptual understanding (posttest scores, argumentation quality), we conducted mediation analyses using the PROCESS macro (Hayes, 2017) in SPSS 25.

5. Results

5.1 Number and Type of Clusters

The *k*-means cluster analysis indicated the optimal cluster solution to consist of three clusters (measured by the largest decrease in within-cluster sum of squares). To validate the cluster solution, we calculated *F*-scores indicating the homogeneity of each cluster (Scheiter et al., 2009). The *F*-score is the ratio of the within-cluster variance to the total variance for each clustering variable ($F < 1$ indicates good homogeneity). Additionally, a discriminant analysis correctly classified 98.2% of the students in the previously identified clusters, indicating a very good fit between the two methods (see Table 2). Large standardized discriminant coefficients indicated that prior knowledge and interest contributed the most to the cluster separation.

Table 2

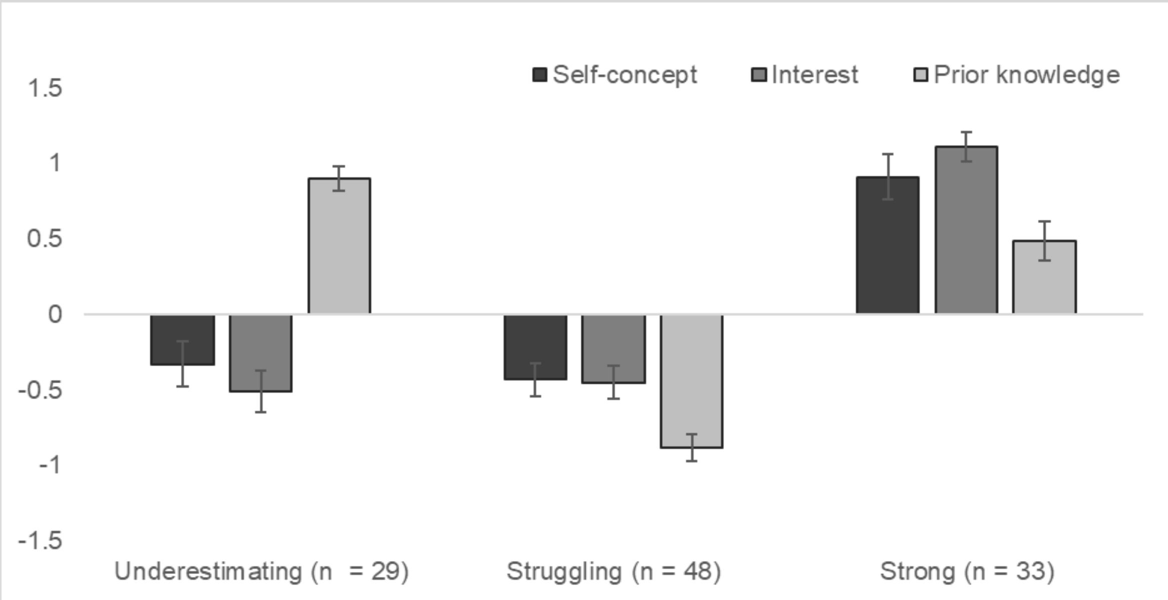
Standardized Homogeneity and Discriminant Indices

Variables	Underestimating (n = 29)	Struggling (n = 48)	Strong (n = 33)	Discriminant indices
Self-concept	0.68	0.59	0.73	.305
Interest	0.56	0.58	0.25	.626
Prior knowledge	0.20	0.35	0.58	.744

The *Underestimating* cluster (n = 29; 26.4%) had slightly below-average academic self-concept, below-average interest, and high prior knowledge, see Figure 4. The *Struggling* cluster (n = 48; 43.6%) scored below-average on all variables. The *Strong* cluster (n = 33; 30%) had high self-concept and interest, and above-average prior knowledge. Students' gender and biology grade were not used for clustering but were analyzed as a function of clusters. The clusters did not differ significantly in gender $X^2(2) = 5.61$, $p = .061$ or in the last biology grade, $F(2, 69.14) = 1.53$, $p = .22$.

Figure 4

Characteristics of the Three Clusters Using Standardized Scores and Standard Error Bars



5.2 Cluster Membership and Experimentation Skills

A sensitivity analysis ($k = 3$, $N = 110$, power = .80) conducted in G*Power 3 (Faul et al., 2007) indicated that the effect sizes would have to be at least medium ($f = 0.30$) to be detected. A Welch one-way ANOVA indicated significant differences between the clusters in the number of controlled experiments, $F(2, 52.78) = 4.35$, $p = .018$, $\omega^2 = .06$ (medium). Post-hoc comparisons identified that, surprisingly, *Struggling* students had a significantly higher proportion of controlled experiments than *Strong* students, $p = .028$, $d = 0.61$. There were no other significant differences (see Table 3). Overall, the proportion of controlled experiments was very high.

5.3 Cluster Membership and Conceptual Understanding

The clusters differed significantly in the conceptual knowledge posttest, $F(2, 109) = 8.81$, $p < .001$, $\eta^2 = .14$ (large). Post-hoc comparisons showed that the *Underestimating* students scored higher than the *Struggling* students, $p = .038$, $d = 0.61$. *Strong* students scored higher than *Struggling* students ($p < .001$, $d = 0.91$), but they did not differ significantly from

Underestimating students, $p = .61$. To determine whether significant posttest differences were not simply a statistical artefact (regression to the mean), we compared the differences in learning gains, measured by subtracting the pretest from the posttest scores. Since the tests had different number of items, we z-standardized the scores, which now reflect learning gains relative to the sample.

Table 3

Means (and Standard Deviations) of Experimentation Skills and Conceptual Understanding

	Underestimating ($n = 29$)	Struggling ($n = 48$)	Strong ($n = 33$)
Proportion of CVS	0.83 (0.26)	0.92 (0.12)	0.83 (0.17)
Correct variable use (absolute numbers)	7.90 (22.65)	8.06 (17.33)	10.85 (6.69)
Incorrect variable use (absolute numbers)	8.48 (22.27)	3.27 (17.29)	10.27 (12.39)
Number of experiments	11.38 (26.66)	8.81 (17.98)	12.06 (10.89)
Conceptual knowledge post test	6.83 (1.73)	5.79 (1.69)	7.39 (1.82)
Learning gains (relative)	-0.75 (0.93)	0.47 (1.14)	-0.03 (1.19)
Claim (0–1)	0.72 (0.46)	0.70 (0.46)	0.81 (0.40)
Evidence (0–1)	0.24 (0.43)	0.18 (0.39)	0.16 (0.37)
Reasoning (0–1)	0.68 (0.48)	0.68 (0.47)	0.74 (0.45)
Argumentation total (0–3) [a]	1.64 (1.15)	1.57 (1.07)	1.71 (1.05)

[a] $n = 25$, $n = 44$, $n = 31$, respectively

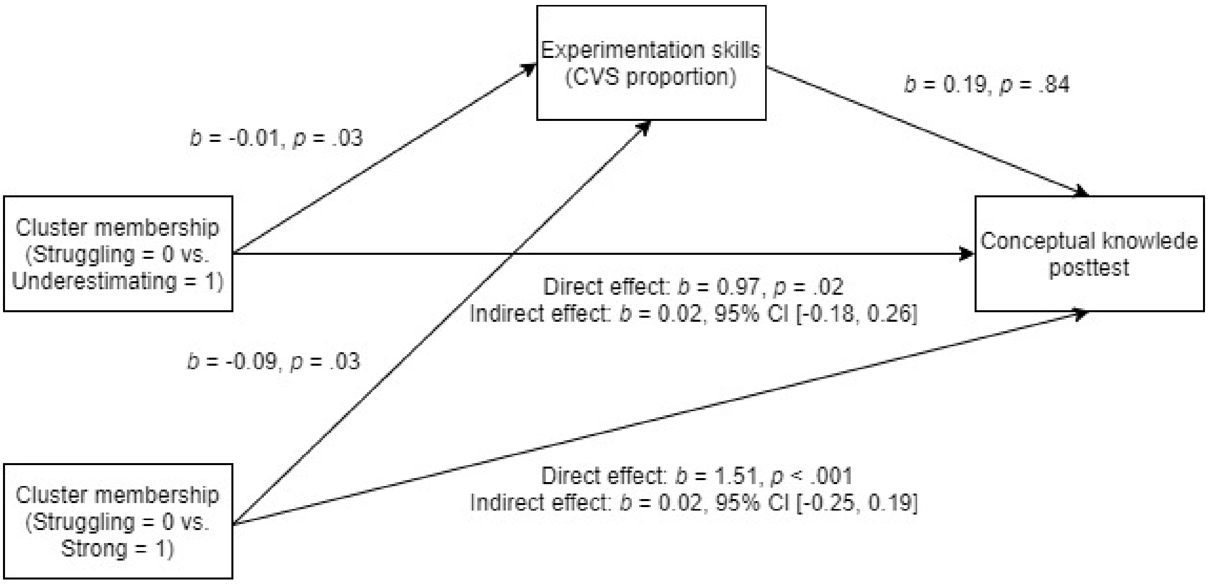
The clusters significantly differed in their learning gains, $F(2, 109) = 11.04$, $p < .001$, $\eta^2 = .17$ (large). Post-hoc comparisons indicated that *Underestimating* students had significantly lower learning gains than *Struggling*, $p < .001$ and *Strong* students, $p = .036$. *Struggling* and *Strong* students did not differ significantly, $p = .14$. The *Struggling* students had the highest learning gains (see Table 3), and prior knowledge was negatively correlated with learning gains ($r = -.60$, $p < .01$), indicating that lower prior knowledge was associated with higher learning gains. The clusters did not differ significantly in argumentation quality. Overall, all students rarely provided evidence in their statements, whereas they more often provided a correct claim and reasoning.

5.4 Mediation Analysis

We conducted mediation analysis to determine if experimentation skills mediate the relationship between individual characteristics (self-concept, interest, and prior knowledge) and conceptual understanding. Since the clusters did not differ in argumentation quality, mediation analyses were performed for the posttest only. The *Struggling* cluster served as the reference category. Therefore, the first set of effects compared the *Underestimating* and the *Struggling* clusters and the second – the *Strong* and the *Struggling* clusters. There were no indirect effects, demonstrated by the 95% bootstrap confidence intervals including zero. Hence, experimentation skills did not mediate the relationship between individual characteristics and conceptual knowledge posttest scores, see Figure 5.

Figure 5

Mediation Analysis with Cluster Membership as Predictor Variable (X), Experimentation Skills as a Mediator (M), and Posttest Scores as Outcome Variable (Y)



6. Discussion

The present study provides a novel application of the person-centered approach in relation to experimentation skills and conceptual understanding during inquiry learning. We identified distinct subgroups of students based on their prior conceptual knowledge, interest, and academic self-concept in biology. We related students' characteristics to their learning processes and learning outcomes during a teacher-led inquiry lesson.

6.1 Types of Student Clusters

We identified one inconsistent (*Underestimating*) and two consistent (*Struggling* and *Strong*) clusters of students based on their cognitive and motivational characteristics. *Underestimating* students had high prior knowledge but low interest and self-concept, whereas *Struggling* students scored low on all variables, and *Strong* students scored above-average on all variables. Intriguingly, we did not identify a group scoring very high on all three variables. *Strong* students had almost twice as high self-concept and interest than prior knowledge. In our sample, prior knowledge was weakly positively correlated with self-concept ($r = .27$, $p = .004$), but not with interest. While a substantial relationship between interest and prior knowledge was identified (Tobias, 1994), a potential explanation is that we assessed general interest in biology whereas the prior knowledge items were topic specific. Topic interest is strongly related to knowledge acquisition and academic achievement (Alexander et al., 1997; Schiefele et al., 1992). Nevertheless, "strong" students identified in prior research do not score equally high on all variables, but show similar patterns (Kosel et al., 2021).

6.2 Conceptual Understanding

Regarding conceptual understanding, *Strong* and *Underestimating* students had significantly higher posttest scores than *Struggling* students. We found no differences in argumentation quality and most students struggled with providing evidence in their answers. Students were not taught to provide high-quality arguments and they require additional instruction and practice to successfully engage in argumentation (Osborne et al., 2004). Thus, both cognitive (prior knowledge, *Underestimating*) and motivational variables (self-concept, interest, *Strong*) were associated with conceptual understanding (posttest).

We additionally computed learning gains from pre- to posttest to ensure that differences in learning outcomes were not simply a result from varying prior knowledge. *Struggling* students showed the highest learning gains, almost twice as high as *Strong*, and four times higher than *Underestimating*. *Underestimating* students, with the highest prior knowledge, had close to zero learning gains between pre- and posttest. This finding suggests an expertise-reversal effect (Kalyuga, 2007), showing that instruction that is highly effective for low-prior knowledge students is ineffective, or even detrimental, for high-prior knowledge students. Similarly, intervention studies in inquiry learning have emphasized the need to tailor guidance to students' prior knowledge. Students with low-intermediate prior knowledge benefitted the most from learning with a less constrained tool which allowed them to design experiments and use the CVS freely (van Riesen et al., 2018, 2019). Our findings indicate that it is important to further investigate the influence of motivational variables during guided inquiry, as interested and confident students might be more autonomous and therefore require less (specific) guidance.

6.3 Experimentation Skills

The clusters differed significantly in the posttest, learning gains, and experimentation skills (only *Struggling* and *Strong*), but not in argumentation quality. Despite conducting the highest proportion of controlled experiments, the *Struggling* students scored lower than *Underestimating* and *Strong* students in the posttest. Nevertheless, the *Struggling* students had the highest learning gains. In the following, we discuss potential explanations of these contradictory findings regarding students' experimentation skills and conceptual understanding.

The fact that the *Struggling* students performed the highest proportion of controlled experiments is surprising, as previous research has demonstrated that low-prior knowledge students apply the CVS less frequently than high-prior knowledge students (Bumbacher et al., 2018; Schauble et al., 1991). However, these studies only considered the effects of prior knowledge while we further considered interest and self-concept. Furthermore, the proportions of controlled experiments for all students were high (between 83% and 92%), suggesting that students often performed the task correctly and followed the structured instruction.

Using structured tasks to reduce the number of actions can support low-prior knowledge students (van Riesen et al., 2018, 2019). On the one hand, our findings attest the benefits of using structured tasks to guide low-prior knowledge students' use of the CVS, as *Struggling* students used the CVS significantly more often than *Strong* students. However, the structured task might have been redundant for students who are knowledgeable, motivated, and confident (*Strong* students). Großmann and Wilde (2019) showed similar findings; however, they only considered students' prior knowledge and not their motivation.

On the other hand, while specific guidance can be beneficial for supporting low-prior knowledge students, too much guidance "inevitably challenges the inherent nature of the inquiry process" (p. 706, Lazonder & Harmsen, 2016) and can hinder deep processing and reflection (Großmann & Wilde, 2019). Providing students with a highly structured task might have impeded them from engaging in thoughtful experimentation. A highly structured task also makes it difficult to determine whether the high proportion of controlled experiments is in fact a result from thoughtful experimentation and understanding or simply following directions. Consequently, our measure of experimentation skills might be unsuitable for measuring the quality of performing on structured task. Process measures can only explain differences in students' actions when students are given enough degrees of freedom. This explanation is further supported by the non-significant mediation of experimentation skills.

To measure intentionality during experimentation, Bumbacher et al. (2018) used the time between experiments. Others have stressed the importance of also measuring cognitive and metacognitive processes during inquiry (van Riesen et al., 2018) using other methods (e.g., think aloud; Omarchevska et al., 2021). For example, it is important to determine the types of learning strategies that students are using (e.g., trial-and-error, means-end analysis) and their meaningfulness. Integrating information about students' cognitive and metacognitive processes during inquiry might further complement the use of CVS as a measure of experimentation skills. It is important to investigate what other reasoning processes students engage in during inquiry and focus on all inquiry phases (e.g., Fischer et al., 2014; Pedaste et al., 2015). For example, combining information on students' evidence evaluation processes after conducting an experiment might be more informative since incorrect data interpretation can override any benefits from using the CVS correctly (Mulder et al., 2014).

6.4 Limitations and Future Directions

We discuss several limitations of the present study. First, because of the smaller sample size, we used cluster analysis instead of mixture modeling. An inherent drawback of cluster analysis is the subjective identification of the cluster number using the scree plot. However, we additionally performed other measures of cluster validation, which indicated that the within-cluster similarity was high and that two different methods led to the same classification in 98.2% of the cases. Despite the sample size, we related students' cluster membership to learning processes and conceptual understanding in an ecologically valid study. Moreover, the three clusters we identified are matching with previous studies using larger sample sizes and advanced methods (LPA; Kosel et al., 2021). Nevertheless, a replication of our findings using larger sample sizes is recommended.

Second, because of the analysis approach and the design of our study, we cannot establish causal conclusions on how students' characteristics affect their experimentation behavior and conceptual understanding. However, person-centered approaches generally do not aim to establish causality; instead, they aim to discover distinct groups of students and relate those to specific outcomes or covariates (Howard & Hoffman, 2018).

7. Conclusion

The present study attests the value of a person-centered approach in relating consistent and inconsistent subgroups of students to learning processes and outcomes in the context of inquiry learning. *Underestimating* students scored as high on the posttest as *Strong* students, indicating that motivational variables also contribute to explaining variability in conceptual understanding. As *Underestimating* students (26.3%) had no learning gains from the inquiry task, it is crucial to design guidance that it is effective for diverse types of students. While this is being considered for prior knowledge (van Riesen et al., 2018; 2019), our findings highlight the importance of considering students' prior interest and academic self-concept when engaging them in inquiry learning.

VII. Study 2: It Takes Two to Tango: How Scientific Reasoning and Self-Regulation Processes Impact Argumentation Quality

1. Abstract

Improving scientific reasoning and argumentation are central aims of science education. Because of their complex nature, self-regulation is important for successful scientific reasoning. This study provides a first attempt to investigate how scientific reasoning and self-regulation processes conjointly impact argumentation quality. In a study with university students ($N = 30$), we used fine-grained process data of scientific reasoning and self-regulation during inquiry learning to investigate how the co-occurrences between scientific reasoning and self-regulation processes are associated with argumentation quality. When modeling the co-occurrence of scientific reasoning and self-regulation processes using epistemic network analysis, differences between students showing either high or low argumentation quality become apparent. Students who showed high argumentation quality engaged in different scientific reasoning processes together more often than students with low argumentation quality, and they made more connections between self-regulation and scientific reasoning processes. These findings offer educational implications for teaching scientific reasoning. Integrating self-regulation and scientific reasoning during instruction could be beneficial for improving scientific reasoning and argumentation.

Published as: 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*. https://doi.org/10.1080/10508406.2021.1966633

2. Introduction

Scientific reasoning and argumentation are crucial facets of scientific literacy and their development has become a fundamental aim of science education at all educational levels (Engelmann et al., 2016; NRC, 2013; OECD, 2013). As a result, science education has moved away from traditional lecture-style pedagogies to more active and learner-centered instructional approaches like inquiry learning (de Jong, 2019; Freeman et al., 2014; Kober, 2015). In these approaches, students need to perform scientific reasoning activities while monitoring and controlling their learning, thereby engaging in self-regulated learning (Greene, Anderson, et al, 2018). Inquiry learning can be more effective for learning than direct instruction (Alfieri et al., 2011; Hmelo-Silver et al., 2007; Lazonder & Harmsen, 2016), but students' success largely depends on their ability to self-regulate their learning (Greene, 2018; Greene, Anderson, et al., 2018; Sinatra & Taasoobshirazi, 2018). Despite the importance of self-regulation for scientific reasoning, research on the interplay of scientific reasoning and self-regulation processes is scarce.

Against this background, we investigated how the co-occurrences of self-regulated learning and scientific reasoning processes relate to argumentation quality. We collected fine-grained process data of self-regulation and scientific reasoning recorded from university students' interaction with a computer-based inquiry learning environment. We identified connections between self-regulated learning and scientific reasoning by focusing on temporal patterns of association between these processes. Epistemic network analysis (ENA, Shaffer, 2017) was used as an innovative method to model the temporal co-occurrence of self-regulation and scientific reasoning processes in relation to argumentation quality.

3. Theoretical Background

3.1 Scientific Reasoning and Argumentation

3.1.1 Scientific Reasoning. Scientific reasoning and argumentation skills are increasingly important for understanding and evaluating the vast amount of scientific knowledge we come across every day (Engelmann et al., 2016; Fischer et al., 2014; NRC, 2013). Scientific reasoning is necessary for understanding how scientific knowledge is created, its associated scientific concepts and methods, and for assessing the validity of scientific findings (Fischer et al., 2014). Consequently, fostering the development of scientific reasoning and argumentation skills has become a fundamental aim of science education (Engelmann et al., 2016; OECD, 2013).

Scientific reasoning has been conceptualized and studied by different research disciplines from developmental psychology (e.g., Kuhn, 2010), cognitive and educational psychology (Klahr & Dunbar, 1988) to science education (Osborne, 2013). Scientific reasoning was conceptualized as a problem-solving process (Klahr & Dunbar, 1988; Zimmerman, 2007) which requires search within a hypothesis space and an experiment space. According to the Scientific Discovery as Dual Search model (Klahr & Dunbar, 1988), three cognitive processes are involved in scientific reasoning – hypothesis generation, experimental design, and evidence evaluation. Kuhn (2010) defined scientific reasoning as an intentional knowledge-seeking process that aims to coordinate theory (or hypothesis) and evidence in four phases – inquiry, analysis, inference, and argument (Kuhn, 2010). One important cognitive strategy regarding coordinating theory and evidence is the control-of-variables strategy (CVS, Chen & Klahr, 1999). The CVS entails holding all variables, except the one being tested, constant during experimentation to produce conclusive results. Using the CVS is essential for differentiating between confounded and unconfounded experiments to draw accurate conclusions from a scientific experiment (Zimmerman et al., 1998).

Lastly, Fischer et al. (2014) provide a framework of scientific reasoning that integrates previous theories and models. It defines scientific reasoning and argumentation as a set of cross-domain skills, comprising eight epistemic activities. Fischer et al. (2014) defined the eight epistemic activities as *problem identification*, *questioning*, *hypothesis generation*, *construction and redesign of artefacts*, *evidence generation*, *evidence evaluation*, *drawing conclusions*, and *communicating and scrutinizing*, see Table 4. The authors emphasize that not all epistemic activities are given equal weight in each scientific domain and that engaging in all epistemic activities is not always necessary for scientific reasoning (Fischer, et al., 2014; Hetmanek et al., 2018).

The scientific reasoning frameworks share the common view that scientific reasoning involves an intentional problem-solving process (Klahr & Dunbar, 1988; Zimmerman, 2007) aiming to co-ordinate theory and evidence (Kuhn, 2010). The epistemic activities shared between the scientific reasoning frameworks discussed above are *hypothesis generation*, *evidence generation*, *evidence evaluation*, and *drawing conclusions*. Hence, in the present paper, we focus on this core set of scientific reasoning processes, which were also relevant to the task students were asked to accomplish.

Table 4

Eight Epistemic Activities and Descriptions Defined by Fischer et al. (2014).

Epistemic activity	Description
Problem identification	Perceiving a mismatch between a problem (from a science, professional, or real-world context) and current explanations, analyzing the situation, and building a problem representation.
Questioning	Identifying one or more questions as the basis for an upcoming reasoning process.
Hypothesis generation	Constructing possible answers to a question (according to scientific standards) based on known models, frameworks, or evidence.
Construction and redesign of artifacts	Creating prototypical artefact (e.g., an engineer building a machine or a teacher constructing a learning environment), testing it, and revising it based on the test.
Evidence generation	Producing evidence following one of several methods. Amongst them are controlled experiments, observational studies, and deductive reasoning based on theory.
Evidence evaluation	Analyzing various forms of evidence in regard to a claim or theory.
Drawing conclusions	Coming to a conclusion by weighing the relevance of different pieces of evidence. Can lead to the revision of an initial claim.
Communicating and scrutinizing	Presenting and discussing the methods and the results of a scientific reasoning process both within a team and a broader community.

Note. From “Beyond intelligence and domain knowledge” by A. Hetmanek, K. Engelmann, A. Opitz, and F. Fischer, 2018, In F. Fischer, C. A. Chinn, K. Engelmann, & J. Osborne (Eds.), *Scientific reasoning and argumentation*, p. 205 (https://doi.org/10.4324/9780203731826-12). Copyright 2018 by Taylor & Francis.

Because students need to engage in these different complex epistemic activities, scientific reasoning and argumentation are often demanding for them (Osborne, 2010). While children and adults can understand the basic difference between hypothetical beliefs and empirical evidence, they have deficits in understanding the mechanisms and the methodology to produce and evaluate evidence-based arguments (Fischer et al., 2014; Koslowski, 2012). Specifically, students often have difficulties developing a hypothesis and designing an experiment to test it (de Jong & van Joolingen, 1998). Still, many adults are struggling with relating different theories, hypotheses, and evidence (Smith & Wenk, 2006). For example, university students demonstrate common false scientific reasoning strategies, related to building and interpreting graphs, and not comparing the results from an experimental group to a control group (Woolley et al., 2018).

The difficulties with scientific reasoning outlined above were found in different domains. For example, students have difficulties understanding and reasoning about genetics (Freidenreich et al., 2011) and even university students can struggle with these concepts (Duncan, 2007; van Mil et al., 2016). Students have issues with understanding the relationship between genes and observed traits and more specifically how the synthesized proteins are involved in the expression of traits (Thörne & Gericke, 2014). The ability to reason about genetic principles is important for general scientific literacy, for example, to understand genetically modified organisms (Haskel-Ittah et al., 2019).

3.1.2 Argumentation. Directly related to scientific reasoning is argumentation, which involves providing evidence-based statements in scientific explanations (Kuhn, 1993; Schwarz, 2009). Argumentation was conceptualized as a part of a broader knowledge-construction process together with scientific reasoning by Fischer et al. (2014). Argumentation in scientific explanations is a consequence of scientific reasoning because it uses evidence to draw conclusions about a scientific process (Engelmann et al., 2016). Hence, argumentation quality is a sensitive indicator of students' ability to reason scientifically.

In this study, to determine argumentation quality, we refer to the claim-evidence-reasoning (CER) model by McNeill et al. (2006), which was designed to support students in constructing scientific explanations. It breaks argumentation down into a claim, evidence, and reasoning. The claim is the conclusion that answers the original question, the evidence is the scientific data that is used to support the claim, and the reasoning is the justification that explains why the data counts as evidence to support the claim, often including scientific principles. According to McNeill et al. (2006), data are considered evidence when they are both appropriate (directly relevant to the claim) and sufficient (e.g., multiple pieces of evidence to support the claim). The CER model is used for teaching students how to incorporate theory and evidence in their arguments to understand scientific principles and the scientific practice of evaluating principles against evidence. In other words, focusing on these aspects of argumentation can support both the acquisition of conceptual knowledge about scientific phenomena and epistemic knowledge (Osborne et al., 2004) about the values of scientific practice, such as using data to justify a claim (Sandoval & Çam, 2011).

In science education, argumentation is important for writing scientific explanations, which can enhance understanding of the scientific content (Zohar & Nemet, 2002). However, similar to scientific reasoning, students struggle with articulating and justifying their claims (Sadler, 2004) and they need practice to improve their argumentation (Osborne et al., 2004). Students also struggle with providing or justifying the evidence for their claims (Bell & Linn, 2000; McNeill, 2011), and they tend to make claims without providing justifications (Fischer et al., 2014). Educated and experienced adults can still struggle with providing evidence for their claims and with differentiating between theory and evidence (Kuhn, 1991).

3.2 Engaging in Scientific Reasoning and Argumentation Successfully

3.2.1 Inquiry Learning. Inquiry learning is one pedagogical method that engages students in scientific reasoning activities, which in turn are assumed to enable them to provide high-quality arguments. A recent trend is to implement inquiry learning in technology-based environments containing virtual experiments (Blake & Scanlon, 2007; Rutten et al., 2012), as they provide a controlled and safe environment for students to practice scientific reasoning skills (de Jong, 1991). Using virtual experiments, students can repeat scientific experiments and investigate a variety of variable configurations. Hence, virtual experiments are a useful addition for eliciting and investigating scientific reasoning processes, because students can test research questions and hypotheses, collect evidence, evaluate evidence, and draw conclusions (de Jong, 1991; van Joolingen et al., 2007; van Joolingen & Zacharia, 2009).

Nevertheless, reasoning scientifically in (virtual) inquiry settings can be difficult for students, not only because they lack scientific reasoning skills, but also metacognitive skills to self-regulate their learning (de Jong & Njoo, 1992; Manlove, et al., 2006). Metacognitive skills are particularly important in such open-ended learning settings which are characterized by non-linear and non-sequential presentation of information and multiple representations of information like graphs, tables, text, and simulations (Azevedo, 2005; Goldman, 2003). To successfully navigate through computer-based learning environments, students are required to exert control of their learning processes (Scheiter & Gerjets, 2007) by selecting relevant learning strategies, information, and variables. Therefore, to comprehensively describe the processes that contribute to students' success (or failure) in providing high-quality arguments, one needs to not only study scientific reasoning processes, but also self-regulatory processes.

3.2.2 Self-Regulated Inquiry Learning. Self-regulated learning is conceptualized as students' use of cognitive and metacognitive strategies to control their learning (Pintrich, 1999). Self-regulated learning models define self-regulated learning not as a trait, but rather as a temporal, dynamic, and adaptive process (Zimmerman, 2013). Students who regulate their cognition engage in processes of planning, monitoring, and evaluation while adopting different learning strategies to achieve their goals (Boekaerts, 1999; Ertmer & Newby, 1996; Schraw, 1998). Students who plan effectively, start by setting goals they want to achieve at the end of the learning task (Winne, 2001). They also plan their time and the learning strategy required for the learning activities. In the next step, students monitor their comprehension and their goal progress, and in case their goal is not met yet, they choose appropriate follow-up strategies and activities (Azevedo et al., 2004; Winne, 2001).

The different models of self-regulated learning (e.g., Boekaerts, 1999; Zimmerman, 2013) have in common that they define self-regulation as a cyclical process comprising different processes, such as monitoring, planning, and regulation of learning activities. The present paper looks in detail at these processes and their interaction with scientific reasoning and cognitive processes, which is why this commonality among models is central to our research. The main difference between these self-regulation models is their comprehensiveness: whereas Winne's (1996) model is focused on cognition, Zimmerman's model is based on social-cognitive theory, and Boekaerts' model (1999) considers how the context influences the type of goals students pursue (Panadero & Alonso-Tapia, 2014). In general, having high self-regulation skills is associated with higher learning gains (Azevedo et al., 2004), while the opposite is true for students struggling to self-regulate (Land & Greene, 2000). This is particularly true for computer-based learning environments, where students often struggle with self-regulation (Azevedo et al., 2010; Bannert, 2009; Bannert et al., 2015).

Students who can self-regulate their learning can benefit more from active learner-centered pedagogies like inquiry learning than their less self-regulated peers (Dent & Koenka, 2016). The complex nature of scientific reasoning, comprising multiple epistemic activities, requires students to actively self-regulate their inquiry learning, to be successful (Manlove et al., 2009b; Reid et al., 2003; White et al., 2009). Planning at the beginning of a scientific reasoning task can be helpful for students to decide which variables they want to manipulate to conduct their experiment. Then, students need to monitor whether they are manipulating the correct variable, or whether they have collected enough data. Students need to recognize if they are not performing these activities correctly (i.e., monitoring) and select a new appropriate learning strategy (i.e., control; White et al., 2009). Therefore, metacognitive factors, like knowledge and regulation of one's cognition, which are important aspects of self-regulated learning (Boekaerts et al., 2000; Pintrich, 2000), can influence knowledge gains from inquiry learning (Chin & Brown, 2000; Kuhn et al., 2000).

Several authors have emphasized the importance of metacognition, which is a central aspect of self-regulated learning, for scientific reasoning (Amsel et al., 2008; Andersen & Garcia-Mila, 2017; Magno, 2011; Manlove et al., 2007; Pedaste et al., 2012). In particular, self-regulation is important for choosing which scientific reasoning strategies to use and when to apply them. However, to our knowledge, a fine-grained analysis of the interaction between self-regulation processes and scientific reasoning activities during inquiry and its impact on argumentation quality is yet missing. One notable exception are the studies by Manlove et al. (2009b), who investigated the effects of regulative support tools on learning outcomes during inquiry learning with online simulations. Their findings showed that using the support tools was positively associated with students' lab report scores and negatively associated with students' model quality. If we assume that using regulative support tools indicates the application of self-regulation processes, then this is evidence of how self-regulation processes matter for the outcomes of scientific inquiry. However, the studies still leave open the question of how scientific reasoning processes and self-regulation processes are used conjointly and spontaneously (i.e., when not prompted through providing support tools). Thus, in the present paper, we argue that the spontaneous co-occurrence of the processes needs to be studied.

3.3 Epistemic Network Analysis as a Path to Analyze Co-Occurrences

Computer-supported learning environments (e.g., virtual experiments) give insight into students' actions during learning by means of log-file data, screen captures, and audio and video recordings (Greene et al., 2021). These data sources incorporate temporal information about the timing and the sequence of learning activities. Next to logfiles, concurrent think aloud protocols during a learning task can provide additional information about students' cognitive and metacognitive processes (Greene, Deekens, et al., 2018).

Still, analyses of these process data have often been limited to what is called *coding-and-counting strategy* (Suthers, 2006) or frequency analysis (Vogel & Weinberger, 2018), where information about the *pattern* of activity over time is lost. This reduces the explanatory power of the analysis and the validity of the conclusions (Reimann, 2009). Consequently, researchers have recommended investigating how variables are connected and the patterns they exhibit over time (Csanadi et al., 2018; Engelmann & Bannert, 2021; Reimann, 2009).

Modeling the sequences of cognitive and metacognitive events can be more meaningful than solely comparing and studying the frequencies of these events in isolation (Bernacki, 2018; Zhou & Winne, 2012).

One novel method for analyzing the temporal relationship between cognitive and metacognitive processes (i.e., scientific reasoning and self-regulated learning) is epistemic network analysis (ENA; Shaffer, 2017). The main rationale behind this approach is that "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). Theoretically, ENA is based on Epistemic Frame Theory (Shaffer, 2007, 2017) which postulates that learning constitutes more than isolated actions during a learning process. Instead, learning is defined as the changes to an individual's epistemic frame which is expressed in interactions (e.g., discourse, exploration of technological artefacts; Shaffer, 2012). The epistemic network represents the set of relationships between the knowledge, skills, and values that a person uses to make decisions or justify their actions (Shaffer, 2017). Shaffer (2017) emphasizes that it is not required to necessarily include all values, skills, and knowledge and that those concepts are just examples of what can be included in an epistemic network. The crucial notion is the *connection* between these concepts.

ENA provides insights into the connection between different processes by measuring their co-occurrence and representing this relationship in dynamic network models (Shaffer et al., 2016). These models portray the structure and strength of the connection between different processes which makes ENA an appropriate method for modeling the structure of connections between self-regulation and scientific reasoning. When we account for the temporal co-occurrence of learning processes, we can observe patterns of learning activities that take place at the same predefined timeframe (e.g., seconds, minutes, learning sessions). Then, we can explore whether these patterns can be observed consistently in our data, and we can conclude that certain learning processes are related to each other, whereas other learning processes are not related (Csanadi et al., 2018). Observation of such patterns is not possible when we only consider the frequencies of single events.

ENA has been used in various disciplines, such as medicine (e.g., communication between health care teams; Sullivan et al., 2018), engineering education (Arastoopour et al., 2016), computer-supported collaborative learning (Csanadi et al., 2018), and diagnostic assessment in medical and teacher education (Bauer et al., 2020). In principle, ENA can be used whenever the research question refers to how groups differ with respect to processes and the way they co-occur. Given the learning sciences' strong reliance on rich process data, ENA appears to be a very promising method for this area of research. Consequently, ENA was proposed as a valuable method for the learning sciences, for example for studying discourse during computer-supported collaborative learning (Csanadi et al., 2018). The advantages of ENA over other modeling techniques is the ability to capture and visualize the temporal patterns of co-occurrence between learning activities in the form of networks. Furthermore, ENA provides a statistical comparison between networks which allows contrasting the co-occurrence of these processes in different groups of students or conditions. Against this backdrop, the present study used ENA to model the temporal co-occurrence of scientific reasoning, self-regulation, and generic cognitive strategies during inquiry learning in relation to argumentation quality.

4. The Present Study

This study investigated how the interaction between scientific reasoning and self-regulation processes impacts argumentation quality, which is important for designing support measures and educational interventions for inquiry learning. Furthermore, by modeling the learning processes of students, next to learning outcomes, we can better explain which aspects of self-regulation and scientific reasoning are associated with higher argumentation quality and inform the design of guidance tools for inquiry learning. We explored how scientific reasoning and self-regulation processes differed between students with high or low argumentation quality in their answers to a research question (the product of the inquiry task). Fine-grained data on scientific reasoning and self-regulation processes came from think aloud protocols and screen recording during a computer-based inquiry task. Guidance is needed for many learners during inquiry (Lazonder & Harmson, 2016); however, because we wanted to observe the full spectrum of processes, only minimal guidance was provided (see method for details). We used ENA (Shaffer, 2017) to model the co-occurrence of scientific reasoning and self-regulation processes in relation to students' argumentation quality. This study had the following research question:

RQ) How do the temporal patterns of association between self-regulation and scientific reasoning processes differ in relation to argumentation quality?

Our underlying rationale was that co-occurrences between scientific reasoning and self-regulation processes would be related to higher argumentation quality. This research question was addressed by modeling the conjoint use of self-regulation, scientific reasoning, and generic cognitive strategy processes using ENA in groups showing high and low argumentation quality. Due to the inherently exploratory nature of ENA, precise hypotheses about specific relationships could not be stated upfront.

5. Method

5.1 Participants and Design

The sample comprised 30 undergraduate students enrolled in different majors (23 females, M_{age} = 23.33 years, *SD* = 2.85) from a university in Southern Germany. Participants were enrolled in both science (*n* = 7; e.g., medicine, biochemistry, nanoscience) and non-science majors (*n* = 23; e.g., economics, law, linguistics, history). Participation in the study was voluntary and written informed consent was obtained prior to participation. Additional informed consent for the audio recording during think aloud was obtained from all participants. The study had a duration of one hour and participants were compensated with 8 € for their time. The study focused on investigating individual differences in students' processes; therefore, no experimental manipulation was used.

5.2 Materials

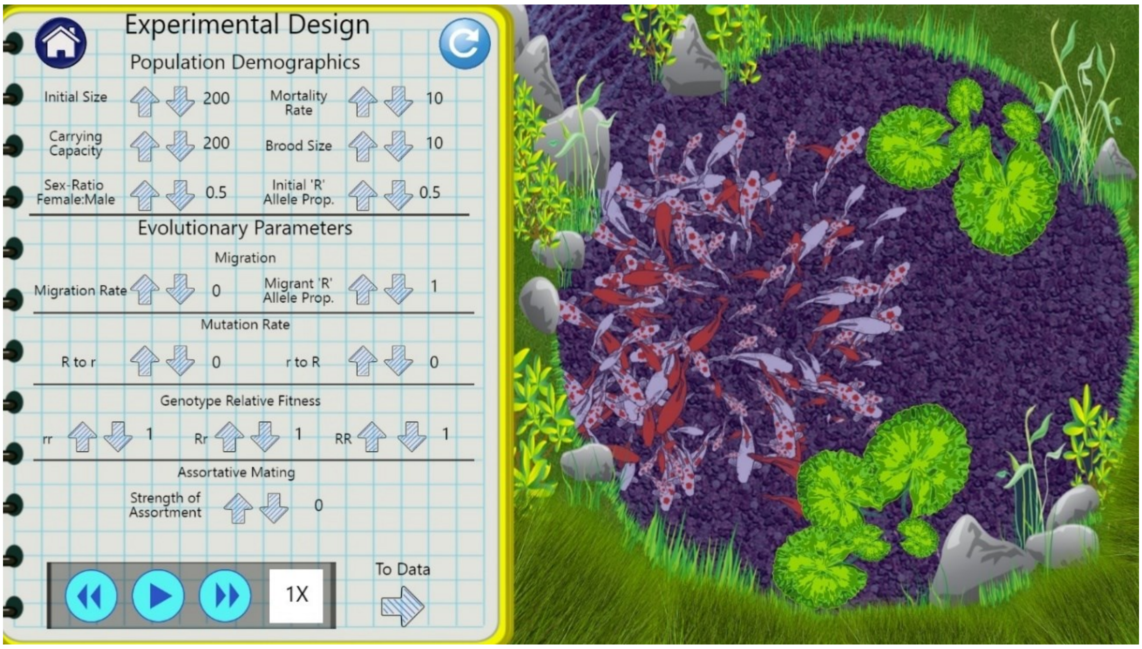
5.2.1 Virtual Experiment. Participants worked with the virtual experiment *PopGen Fishpond* (http://virtualbiologylab.org/ModelsHTML5/PopGenFishbowl/PopGenFishbowl.html), in which they could conduct experiments on the topic of population genetics. In the virtual experiment, participants could manipulate variables related to the population demographics (e.g., population size, male-female ratio, mortality rate), evolutionary parameters (e.g., mutation and migration rate, genotype relative fitness), see Figure 6a. By manipulating different variables, participants could investigate specific variables' influence on the population size, the allele proportions, and the genotype proportions. Participants could analyze the data they collected in a table, a line graph, and a bar graph (see Figure 6b).

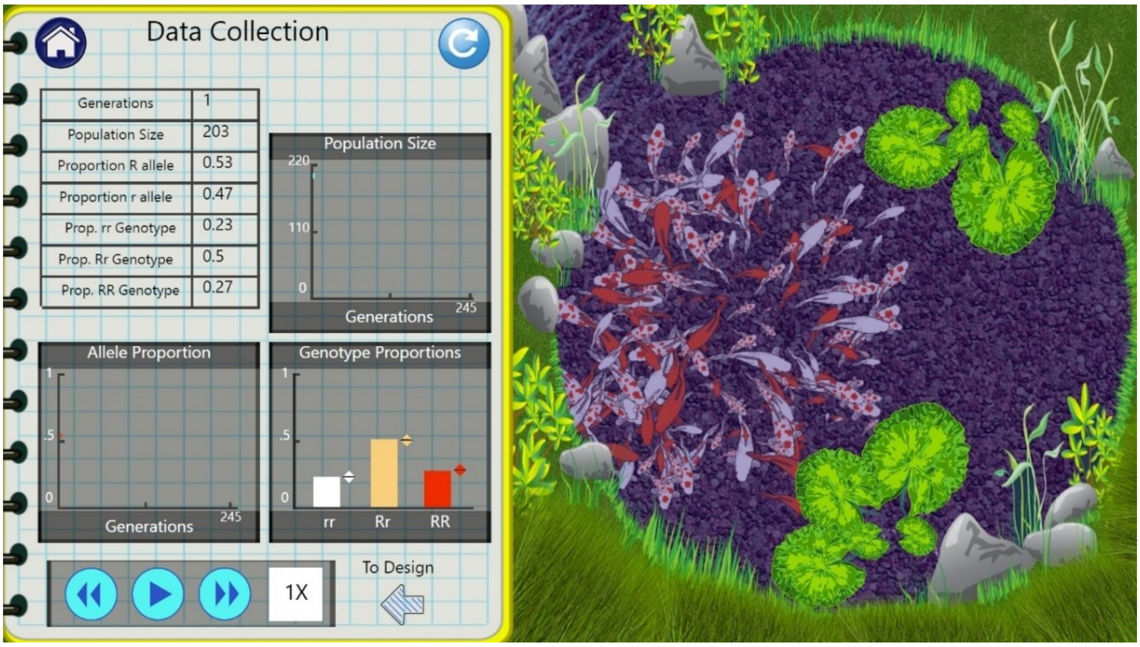
5.2.2 Instruction. Participants read a short text (1 page, 475 words, 1 picture) about natural selection which briefly explained the concepts of genotype, phenotype, natural selection, alleles, and relative genotype fitness. The instruction provided participants with the prerequisite conceptual knowledge necessary to work with the virtual experiment. Since all students had covered population topics during their secondary education, this instruction mainly served to re-activate their knowledge.

5.2.3 Inquiry Learning Task. Participants were asked to use the virtual experiment to verify a statement made by a fictitious other person. The statement was "*In dominant alleles, the reduction in relative genotype fitness has no effect on the dominant genotype proportion over time.*" and it was incorrect. To provide minimal guidance and ensure that the students would have an idea on how to get started with their exploration, they were given specific values to use for genotype relative fitness (0.2, 0.5, and 1) and they were asked to collect data from 200 generations of fish reproduction because this is when the results of the virtual experiment stabilized. Then, they had to interpret the genotype proportion of the dominant genotype (RR) and compare the proportions for the three values of genotype relative fitness. Participants were asked to reset the virtual experiment before conducting a new experiment. In this way, we again provided limited guidance during the inquiry to not completely overwhelm the participants. At the end of the task, participants wrote a short text, in which they explained whether the original statement was correct, and they were asked to provide evidence supporting their conclusion.

Figure 6

PopGen Fishpond Virtual Experiment, with an Experimental Design Pane (a) and Data Collection Pane (b)





Note. Copyright (2016) Virtual Biology Lab.

5.3 Measures

5.3.1 Domain-Specific Academic Self-Concept. Academic self-concept is important for learning since it can affect achievement and interest (Möller et al., 2020; Trautwein & Möller, 2016). Participants rated their domain-specific academic self-concept (6 items, Cronbach's α = .92; Grüß-Niehaus, 2010; Schanze, 2002) on a Likert scale ranging from 1 (*I do not agree at all*) to 4 (*I completely agree*). The items of the scale were adapted to the domain of biology. An example item of the scale is 'I quickly learn new material in biology'.

5.3.2 Domain-Specific Interest. Interest is an important factor for learning and motivation since both children and adults who are interested in a certain activity pay more attention to the activity and they are more persistent and engaged (Hidi, 2001). Domain-specific interest was assessed using a 5-item scale (Cronbach's α = .92; Wilde et al., 2009) which we adapted to the domain of biology. Participants rated their interest in biology on a Likert scale ranging from 1 (*I do not agree at all*) to 4 (*I completely agree*). An example item of the interest scale is 'I am interested in learning something new in biology'.

5.3.3 Conceptual Prerequisite Knowledge. To assess conceptual prerequisite knowledge in genetics (e.g., "How do geneticists represent two recessive alleles?"), we developed four multiple-choice items (Cronbach's α = .63). Due to the low internal consistency of the scale, it was omitted from further analyses.

5.3.4 Scientific Reasoning Ability. Scientific reasoning ability was assessed using six items in the pre-test (Cronbach's α = .08) and 11 different items in the post-test (Cronbach's α = .57). Both tests combined items from the PISA released science items (OECD, 2015) and items developed by Koenen (2014). Because of the low internal consistency of the scales, they were omitted from further analyses.

5.3.5 Scientific Reasoning and Self-Regulation Processes. Think aloud protocols and screen captures were used to obtain the processes of scientific reasoning and self-regulation. Descriptions and examples of the major and sub-categories are provided in Table 5. The coding scheme for scientific reasoning processes was based on the scientific reasoning framework proposed by Fischer et al. (2014). The coding scheme focused on the following epistemic activities: problem identification, hypothesis generation, evidence generation, evidence evaluation, drawing conclusions. We focused our analysis on these epistemic activities, as they could be observed during computer-based inquiry learning. We coded self-regulation processes from the think aloud protocols and we focused on the processes of planning, monitoring, and control. From the think aloud protocols, we additionally coded participants' use of generic cognitive strategies, namely activation of prior knowledge and self-explanation, as these strategies are relevant to learning in general.

We used Mangold INTERACT® (Mangold International GmbH, Arnstorf, Germany; version 9.0.7) software to simultaneously code the screen captures (video) and the think aloud protocols (audio). Initial coding of 20% of the data yielded insufficient inter-rater reliability; therefore, the two raters discussed their disagreements and coded an additional 20% of the data. We used benchmarks for interpretation indicating substantial agreement for kappa values between .61 and .80, and perfect agreement for values between .81 and 1.00 (Landis & Koch, 1977). On average, the inter-rater reliability was high (κ = .84) and specifically, inter-rater reliability was perfect for scientific reasoning processes (κ = .86), and for cognitive and metacognitive processes (κ = .81). After resolving all disagreements on the double-coded data, one of the raters coded the remaining data.

5.3.6 Argumentation Quality. To measure the argumentation quality of students' answers, we adapted the scoring rubric from McNeill et al. (2006) to the context of our study, see Table 6. The original scoring rubric had three components (claim, evidence, reasoning) and each component was given a score between 0-2. Two raters independently scored 20% of the data (n = 6). First, the quality of the claim (κ = .77; percent agreement = 88.3), the evidence (κ

= .74; percent agreement = 88.3), and the reasoning (κ = .54; percent agreement = 66.7) of each answer were scored, adding up to a maximum score of 6. After resolving disagreements, one of the raters coded the remaining answers.

5.4 Procedure

Data were collected in individual self-paced sessions lasting 1 hour, in which participants were seated in front of a laptop in a quiet room. At the beginning of the session, participants were informed about the purpose of the study. We obtained written informed consent for the use of voice recording and participation in the study from all participants. The study was computer-based, and all learning materials were embedded in the online inquiry learning space *Graasp* (http://graasp.eu).

First, participants completed the pretest in Qualtrics (Qualtrics, Provo, UT), and then they started working in the online inquiry learning space. On the first page in Graasp, participants received general information about the experiment, instructions, and two exercises to practice thinking aloud. In the first think aloud exercise, participants were asked to multiply 24 by 36 and verbalize their thoughts during the multiplication. In the second exercise, they were asked to answer the following question: 'Your sister's husband's son is your children's cousin. How is he related to your brother?' and verbalize their thoughts. On the next page in Graasp, participants were provided with background information about natural selection and genetics. Participants then watched a short video tutorial (1 min 45 s) which explained the functions and the interface of the virtual experiment.

Subsequently, participants started solving the inquiry task using the virtual experiment during which they were asked to think aloud. To correctly solve the task, participants had to conduct three experiments, in which they had to vary the value for the genotype relative fitness of the dominant allele in each experiment. At the end of the inquiry task, participants answered the original research question and wrote down a short text from which we coded argumentation quality. We recorded screen captures of participants' interaction with the virtual experiment using Camtasia Studio 8 to analyze process measures of scientific reasoning. Camtasia Studio also simultaneously recorded the audio from the think aloud protocols. Last, participants completed the post-test in Qualtrics.

5.5 Data Analysis

To model the temporal co-occurrences between the processes of self-regulation, scientific reasoning, and generic cognitive strategies in relation to argumentation quality, we used ENA (Shaffer, 2017; Shaffer et al., 2016; Shaffer & Ruis, 2017). ENA can identify associations between codes or variables (e.g., scientific reasoning and self-regulation processes) by quantifying their co-occurrence within a *conversation* (Shaffer et al., 2016). In epistemic network analysis, a conversation is defined as “a set of lines of data that can be related to one another” (Shaffer, 2017, p. 152). By modeling the connections between codes, ENA produces a set of graphs that capture their relationship: 1) a plotted point, which shows the location of a specific unit’s network, and 2) a weighted network graph.

Table 5

Coding Scheme Used for Assessing Scientific Reasoning, Self-Regulation, and Generic Cognitive Strategies

Major category	Sub-category	Source	Definition and example
Scientific Reasoning			
	Problem Identification	TA	Problem identification is concerned with analyzing the situation and building a problem representation. *"I need to check what happens when I decrease the genotype fitness"*
	Hypothesis Generation	TA	Possible answers to the research question are identified and hypotheses are generated. *"So, I expect that increasing the genotype fitness will decrease the genotype proportion."*
	Evidence Generation		Collecting data using controlled experiments.
	Establishing a baseline	Video + TA	Participant sets all values for genotype fitness to 0.5 or to another constant value.
	Variable manipulation	Video	Participant uses CVS and manipulates the correct, incorrect, or irrelevant variable.
	Value usage	Video	The participant uses the values 0.2, 0.5, 1 or other values.
	Evidence Evaluation	Video + TA	The generated evidence is analyzed with regard to the original claim.
	Graph evaluation	Video + TA	The participant interprets the correct or incorrect graph.
	Variable evaluation	Video + TA	The participant interprets the correct or incorrect variable in the graph.
	Interpretation	Video + TA	The interpretation of the graph is correct or incorrect.
	Drawing Conclusions	TA	Weighing different pieces of evidence to come to a conclusion regarding the original claim. *"I think the original statement is incorrect."* (correct conclusion)
Self-Regulated Learning			
	Planning	TA	Specifying sequences of activities for achieving goals. *"Now I will set it [relative genotype fitness] to 1 and see if I can notice any difference."*
	Monitoring (positive, negative)	TA	Being aware of own comprehension and task performance. *"Now I understand how this works."* (positive monitoring) *"I don't know how to conduct an experiment."* (negative monitoring)
	Control	TA	Attempt to guide own cognition, behavior, attention, and motivation based on monitoring. *"That would now confirm the hypothesis, but I will continue experimenting."*
Generic Cognitive Strategies			
	Activation of Prior Knowledge	TA	Participant is referring to prior knowledge (from instruction, school, university). *"I read in the text that rr is the recessive allele."*
	Self-explanation	TA	Participant is explaining what they are observing, and they are providing deeper elaborations. *"So, if I decrease the fitness, then the frequency also decreases…"*

Note. TA = Think aloud, Video = Screen capture

Table 6

Coding Scheme Used for Assessing Argumentation Quality

	Level		
Component	*0*	*1*	*2*
Claim			
Criteria	Does not make a claim or makes an inaccurate claim.	Makes an accurate but incomplete claim.	Makes an accurate and complete claim.
Example from data	“The statement is correct”, “True”, “Yes”, “Correct”	“True in the case that both alleles are [dominant] but not when only one is dominant.”	“The statement is not true …”
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	“Reducing the genotype fitness doesn’t affect the genotype proportion.” (no evidence)	“The dominant part is reduced.” (insufficient)	“…because the genotype proportion did change over time by reducing the fitness.”
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	“The statement is wrong because all genotypes are 0 besides RR. Therefore, there is clearly an influence.”	“The lower the fitness, the faster the extinction.”	“The lower the relative fitness, the lower the genotype proportion.”

Note. Adapted from “Supporting Students’ Construction of Scientific Explanations by Fading Scaffolds in Instructional Materials” by K. L. McNeill, D. J. Lizotte, J. Krajcik, and R.W. Marx, 2006, *The Journal of the Learning Sciences*, *15*, p. 189. Copyright 2006 by Lawrence Erlbaum Associates.

Epistemic network analysis uses a *moving stanza window* to construct a network model for each line of data, which shows how each code is connected to other codes temporally (Siebert-Evenstone et al., 2017). *Stanza* is defined as "a set of lines in a single conversation that are within the same relevant context, and therefore are related to one another" (Shaffer, 2017, p. 154). A moving stanza window is a fixed number of lines that are used to analyze the recent temporal context of a code with a fixed number of codes preceding it. The window is "moving" since a summary value is computed for each code, based on the previous codes, instead of using a summary value for all codes during an activity (Siebert-Evenstone et al., 2017). This allows us to model not only direct connections between processes but also relationships between preceding or following processes.

The associations between codes are derived based on their co-occurrences within stanzas. Then, using a binary summation, networks are aggregated where each line reflects the presence or absence of the co-occurrence of each pair of codes. To account for the fact that different units of analysis (participants) have different amounts of codes, the networks were normalized before they were subjected to a dimensional reduction, which accounts for the fact that different units of analysis may have different amounts of coded lines in the data. For dimensional reduction, epistemic network analysis uses singular-value decomposition (svd), which produces orthogonal dimensions that maximize the variance explained by each dimension (Shaffer et al., 2016).

The nodes in the network graph are fixed and their position is determined by an optimization routing that minimizes the difference between the plotted points and their corresponding network centroids. The centroid is the mean position of all points in the network in all coordinate directions. In this way, ENA is useful for comparing units of analysis (e.g., participants in different groups) by looking at their position in the plotted point, their individual networks, and the weighted networks. Furthermore, networks can be compared using network difference graphs by subtracting the weights of each connection from one network to the corresponding connection of the other network. The validity of the resulting networks can be assessed using the strength of correlation between the centroids and the projected points in the model using Spearman's rho and Pearson's r, which determine the model fit (Shaffer et al., 2016). The reliability of the model can be assessed by going back to the original data to determine whether the relationships obtained in the ENA are reflected in the original data (closing the interpretative loop; Shaffer, 2017). This can be easily done in ENA since we can click on a specific relationship and obtain the portions of the dataset responsible for the relationship.

In this study, we used the ENA Web Tool (version 1.6.0) (Marquart et al., 2018) to compare the networks of participants with low argumentation quality and high argumentation quality. We defined the units of analysis as all lines of data associated with a specific value for argumentation (0 = low argumentation quality, 1 = high argumentation quality) and subsetted it by participant. To illustrate, one unit consisted of all the codes for a participant who belonged to the high argumentation quality group. Because we used data from one short inquiry task, we defined *conversation* as all lines of codes from the task. We used a stanza window of 3 lines, meaning that we modeled the co-occurrence of a specific code with the two codes preceding it. Our model included the following codes: *problem identification (correct) hypothesis generation (correct and incorrect)*, *evidence generation (correct and incorrect)*, *evidence evaluation (correct and incorrect)*, *drawing conclusions (correct and incorrect)*, *control*, *monitoring (positive and negative)*, *activation of prior knowledge* and *self-explanations (correct and incorrect)*. *Problem identification (incorrect)* and *planning* were not connected to other codes and were excluded from the analysis.

6. Results

On average, participants in our sample had a mean domain-specific academic self-concept of 2.36 (*SD* = 0.59) and domain-specific interest of 2.55 (*SD* = 0.67) from a maximum of 4.

6.1 Epistemic Network Analysis (ENA)

ENA allows for statistical comparison between networks of students in different groups, or between individual students. ENA creates several graphs – an individual plotted point graph, group networks graph, and a network difference graph. The difference network (Figure 7) shows qualitative differences between the networks, illustrated by stronger or weaker connections between codes in different groups. The size of the node indicates the importance of each node for the groups (i.e., larger nodes are more important). The thickness of the line is an indicator of the proportion of stanzas in which the two codes co-occurred (Shaffer & Ruis, 2017). In other words, the thickness of the line connecting two nodes represents their frequency of co-occurrence, with thicker lines portraying stronger frequency of co-occurrence. The squares represent the centroids of each network and their confidence intervals. In our case, the red (left) square is the centroid of the low argumentation quality group, and the blue (right) square is the centroid of the high argumentation quality group.

6.1.1 K-Means Cluster Analysis. We were interested in whether the temporal patterns of co-occurrences between scientific reasoning, self-regulation, and generic cognitive strategies differed between participants with high and low argumentation quality. To answer this question using ENA, we created two groups based on their argumentation quality scores using k-means cluster analysis (Everitt et al., 2011). We used participants' scores for claim, evidence, and reasoning (see Table 6) as input variables for the cluster separation. Because we were interested in differences between high and low argumentation quality, we set the desired cluster number to 2. This resulted in two groups – one group with high argumentation quality (n = 15) and one group with low argumentation quality (n = 15). Descriptive statistics of both groups are provided in Table 7.

Table 7

Mean Scores (and SDs) of the Two Groups Resulting From K-means Cluster Analysis (N = 30)

	Argumentation quality	
	Low (n = 15)	High (n = 15)
Claim	0.13 (0.52)	1.13 (0.99)
Evidence	0.40 (0.51)	1.80 (0.41)
Reasoning	0.47 (0.52)	1.67 (0.49)

Note. Claim, evidence, and reasoning scores ranged between 0-2.

First, we correlated argumentation quality with the control variables (prerequisite conceptual knowledge, domain-specific interest, and self-concept) to ensure that prior group differences did not account for some of the variability in argumentation quality. We compared the groups using Kendall's Tau-b, which is the appropriate test for ordinal variables (argumentation quality). The tests revealed no significant association between argumentation quality and domain-specific academic self-concept, τ_b = -.10, p = .54, which indicates no differences in self-concept. Argumentation quality had a moderate negative association with domain-specific interest (τ_b = -.32, p = .044), indicating that low argumentation quality was associated with higher interest.

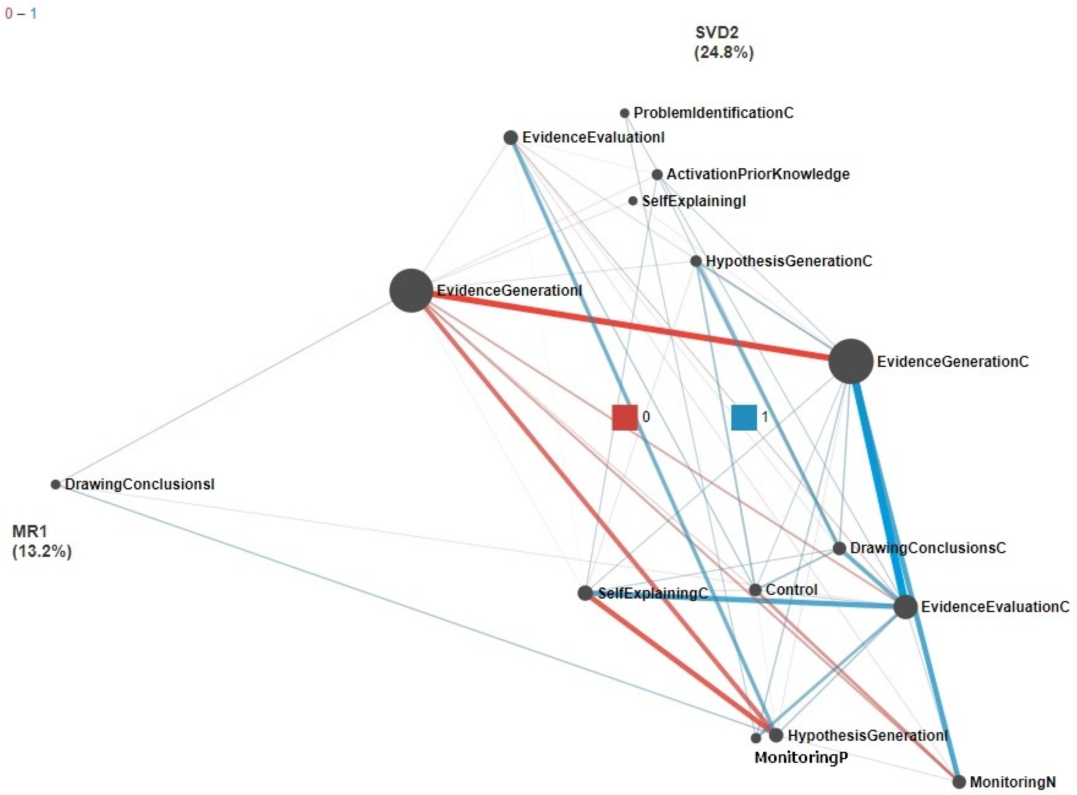
6.1.2 Group Comparison. We compared the centroids of the two networks along the X and Y axes. The first dimension of the rotation went through the means (i.e., means rotation) to find the maximum difference between students who showed high and low argumentation quality (Shaffer, 2017). Because we did a means rotation on X-axis, we forced the two groups to have the same value (0) on the Y-axis. In this way, the variance on the X-axis is determined by participants' group membership (i.e., high or low argumentation quality). A Mann-Whitney test showed that, along the X-axis (dimension 1 after means rotation), the centroid value of the low argumentation quality group (Mdn = -0.28, n = 15) was significantly different from the centroid value of the high argumentation quality group (Mdn = 0.19, n = 15), U = 34.00, p < .001, d = 1.96). This indicates that the two networks differed significantly from each other on the X-axis, which explained 13.2% of the variance in the graph. The two groups did not differ significantly on the Y-axis (U = 104.00, p = .74, d = 0.16), which is a result of the means rotation. The Y-axis explained 24.8% of the variance in the graph. The model had good model fit, indicated by the co-registration correlations between the centroids and the projected points (dimension 1: r = .91, r_s = .90; dimension 2: r = .93, r_s = .92).

Significant differences in centroids indicate that the mean networks of participants in both groups are different. In turn, this implies that the structure of connections within the two groups' networks differ regarding the way that scientific reasoning and self-regulatory processes co-occur. The subtracted network graph (Figure 7) illustrates the differences in strength between individual connections in the two groups' networks. Stronger co-occurrences in the low and the high argumentation quality groups are depicted in red and blue lines, respectively. The thickness of the lines indicates how often two codes were co-occurring and thicker lines correspond to higher frequencies of co-occurrence. The high argumentation quality group's network (blue) is denser, which is illustrated by a higher number of connections between codes, and more specifically between correctly engaging in various scientific reasoning processes. In contrast, in the low argumentation quality group's network (red), many of the stronger connections were made between incorrect evidence generation and other scientific reasoning processes (e.g., hypothesis generation).

On the left side of Figure 7, we observe several connections between incorrectly generating evidence (EvidenceGenerationI) and correctly generating evidence (EvidenceGenerationC) and incorrectly generating a hypothesis (HypothesisGenerationI), and between incorrect hypothesis generation and correct self-explaining (SelfExplainingC). Since these connections co-occurred more frequently in the low argumentation group, they are displayed in red. The position of the low argumentation quality group's centroid to the left is determined by these connections.

Figure 7

A Comparison Plot Showing Differences in the Epistemic Networks of the High Argumentation Quality Group (blue, right) and the Low Argumentation Quality Group (red, left)



Note. For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.

On the right side of Figure 7, several blue connections can be observed. For example, correct evidence generation (EvidenceGenerationC) and evaluation (EvidenceEvaluationC), correct evidence generation and monitoring (MonitoringN), and correct evidence evaluation and drawing conclusion (DrawingConclusionsC). This indicates that participants who showed high argumentation quality were not performing scientific reasoning activities in isolation; rather, these activities were performed more often together and correctly in comparison with the low argumentation quality group. Moreover, we observe stronger co-occurrences between monitoring and scientific reasoning processes. Next, we discuss the specific relationships we found in the two groups, and we illustrate them using examples from the think aloud protocols and the screen captures. In this way, we are going back to our original data to investigate whether our conclusions are grounded (Shaffer, 2017) and we attempt to bridge quantitative and qualitative findings by providing narrative explanations of the connections observed in the networks.

6.1.3 High Argumentation Quality Group. In the high argumentation quality group, we observe more connections between scientific reasoning processes overall. Moreover, the scientific reasoning activities were performed together more often correctly than in the low argumentation quality group. For example, in this network, correct evidence generation (EvidenceGenerationC) is strongly connected with correct evidence evaluation (EvidenceEvaluationC) and drawing conclusions correctly (DrawingConclusionsC). To illustrate, towards the end of their inquiry, one participant concluded (drawing conclusions correctly) and then evaluated the evidence (correctly) by looking at the graphs in the simulation. The participant said '*It already appears that this statement is not correct. Within the first two runs, the dominants were very different in their numbers. I will have another look at the red fish. Now there are fewer white fish and more red fish, but the difference is not as distinct as before.*' In this example, we also observe that this participant was evaluating the evidence from three different experiments while drawing their conclusions.

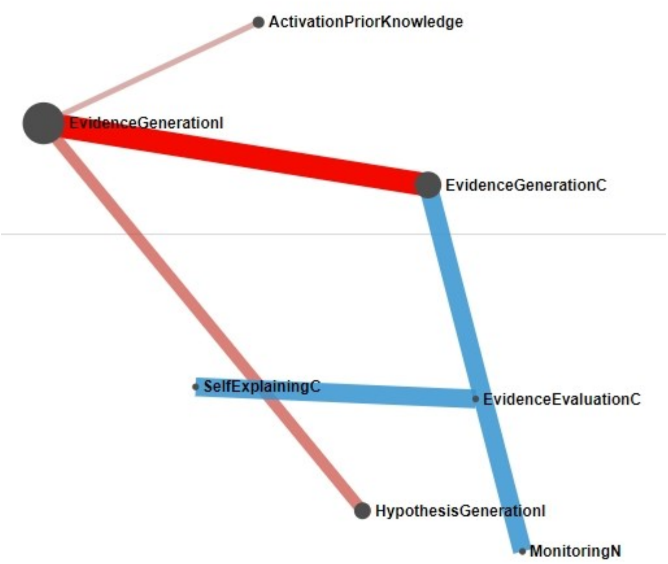
Monitoring (negative) was also more strongly connected with correct evidence generation than in the low argumentation quality group. For example, one participant started the task by manipulating the correct variable and use the control-of-variables strategy (correct evidence generation). Then, she said '*I'm really bad at biology.*' (negative monitoring). She continued to conduct the experiment correctly while monitoring (negative) '*I am not really sure if what I do is right.*' At the end of the experiment she was also self-explaining while looking at her results (evidence evaluation) '*Alright, we are at 200 [generations] again. And the value for the dominant genotype is 0.02. So, I don't have to do the last experiment, because the hypothesis is incorrect.*'. Her epistemic network is shown in Figure 8 (right, blue), in comparison with a participant from the low argumentation quality group (left, red). In Figure 8, we observe the connections that we described qualitatively – thick blue lines between monitoring and evidence generation, and between self-explaining and evidence evaluation.

Furthermore, we observe stronger connections between correct self-explaining and correct evidence evaluation. For example, while looking at the graph showing the result of his experiment, the participant said '*Now I definitely see that something changes. Now it is making sense. If the dominant [gene] is fitter than the other [gene], it should spread more easily. Now*

I see that all my fish are becoming red. So, I understand it now.' This illustrates that this participant was engaging in deeper processing of the evidence during evidence evaluation by self-explaining his observations. Furthermore, we observe three instances of monitoring in this short excerpt of the think aloud protocol – the participant is aware that he has understood the concept and now “it makes sense”. To illustrate this finding further we provide example networks from participants in this group together with their problem-solving behavior and think aloud protocols.

Figure 8

The Subtracted Network of Two Exemplary Participants Who Showed High Argumentation Quality (blue) and Low Argumentation Quality (red)



Note. For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.

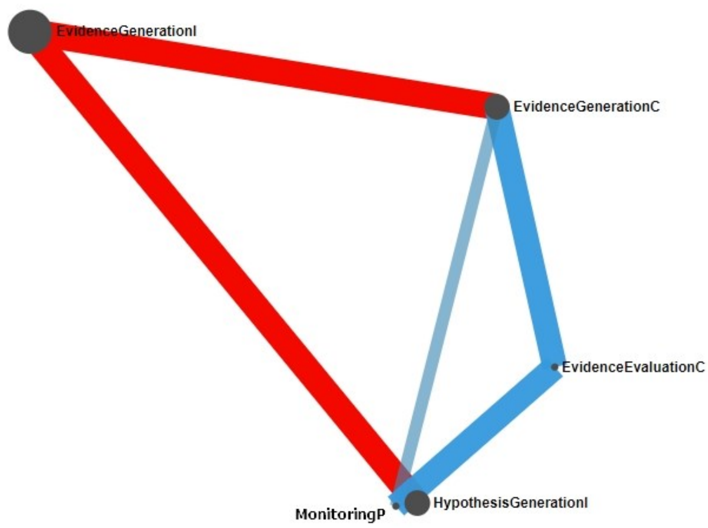
The participant started the task by repeating the statement she was asked to test, and she identified which are the dominant alleles. First, she set the value for genotype relative fitness of the dominant allele to 0.2, as required by the task. She used the control of variables strategy and manipulated only the correct variable, which are both indicators of correct *evidence generation*. Then, she was collecting data and was *monitoring* (positive) her understanding of the task: ‘*Okay, I think I understand it now.*’ After collecting enough data, she saw that the fish with the dominant genotype had disappeared, which disconfirmed the original statement. She expressed her surprise by saying '*[Genotype] proportion is 0, huh?*' (correct *evidence evaluation*). Then, she went back to the virtual experiment and changed the genotype relative fitness of the dominant allele to 0.5 and used the control of variables strategy again (correct *evidence generation*). At the same time, she was *monitoring* again by asking herself: '*Does that work? Yes.*'. She realized that she was still collecting data from her previous experiment and asked the experimenter '*Ahhm, do I have to restart again?*'. She conducted two experiments by setting the dominant genotype relative fitness to 0.5 and 1. Last, she evaluated the evidence which showed that the genotype proportion of the dominant allele was small in all the experiments she conducted. This disconfirmed the original statement and was the correct conclusion.

She wrote the following answer at the end of the task: 'Reduction of the relative fitness of dominant alleles does, in fact, have an influence on their genotype [proportion]. The lower the relative [genotype] fitness, the lower the genotype proportion.' This answer is an example that received the maximum score (6) for argumentation. In the epistemic network of this participant (Figure 9, right), we can see that monitoring was central to her problem-solving process. In Figure 9, we see connections between monitoring (MonitoringP) and correct evidence generation (EvidenceGenerationC) and correct evidence evaluation (EvidenceEvaluationC). These connections were explained qualitatively, and they indicate that metacognitive monitoring during performing scientific reasoning activities was associated with high argumentation quality.

6.1.4 Low Argumentation Quality Group. In the low argumentation quality group, there were fewer stronger connections between the codes overall. We observe that incorrect evidence generation had the most connections to other processes. We observe also a strong relationship between generating evidence correctly and incorrectly. When looking at the screen captures, we can see that this connection is mainly due to violations of the control of variables strategy. This is illustrated by participants first manipulating the correct variable (EvidenceGenerationC), but then also manipulating irrelevant variables (EvidenceGenerationI) immediately after, thereby violating the control of variables strategy. Most often, participants were manipulating the genotype relative fitness for all genotypes at the same time, instead of only manipulating the one for dominant genotypes.

Figure 9

The Subtracted Network of Another Two Participants Who Showed High Argumentation Quality (blue) and Low Argumentation Quality (red)



Note. For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.

In this group, we also found a stronger relationship between generating an incorrect hypothesis and correctly self-explaining. To illustrate this relationship, we went back to the screen captures and think aloud protocols and we provide narrative examples of the participants' problem-solving behavior. During the experiment, one participant stated an incorrect hypothesis '*Now the other two values should increase.*' After observing that there were only red fish left in the pond, the participant explained to herself that her hypothesis was impossible by saying '*Okay, of course the values can't increase, when I already killed the other two types. If there aren't any [recessive genotypes], there can't be new [recessive genotypes].*' Here we can see that self-explaining was beneficial to the problem-solving process and the participant realized that their hypothesis was incorrect.

Last, we found a stronger relationship between incorrect hypothesis generation and incorrect evidence generation. For example, one participant stated their hypothesis '*I think this [genotype relative fitness] really doesn't have an effect.*' Then, the participant manipulated the genotype relative fitness for the wrong genotype (incorrect evidence generation). Only after collecting data for a couple of seconds, the participant concluded that '*Yeah, I really think this doesn't have an effect.*' This participant's epistemic network is shown in Figure 8 (left, red). In Figure 8, these relationships are presented on the left side – incorrect hypothesis generation (HypothesisGenerationI) and incorrect evidence generation (EvidenceGenerationI). To illustrate these findings further, we next discuss the problem-solving process of another participant from the low argumentation group which further demonstrates the relationship between correct and incorrect evidence generation and incorrect hypothesis generation and incorrect evidence generation.

This participant started directly with changing variables regarding both recessive and dominant genotypes. In particular, she changed two irrelevant variables (mutation rates) and then she changed the values for genotype relative fitness for all three genotypes. She said '*Okay, I will change the recessive and also the dominant.*' (incorrect hypothesis generation). One of the values she changed was for the dominant genotype, which was correct evidence generation. However, she manipulated five variables at once, thereby violating the control of variables strategy (incorrect *evidence generation*). Then, without going to the data pane and collecting enough data, she interpreted the results by only looking at the color of the fish in the pond. She manipulated another three irrelevant variables (initial population size, migration rate, and initial allele proportion) and collected more data (incorrect *evidence generation*). She concluded that '*Okay, dominant are the red ones, nothing changes. They always remain the same. The assumption seems to be correct.*' Then she went back to the virtual experiment and changed three more irrelevant variables again (mutation, migration, and migration allele proportion) without using the control of variables strategy (incorrect *evidence generation*). Without waiting to collect more data, she concluded '*I can now confirm that the assumption seems to be correct.*' Her conclusion was incorrect. Here we can observe that starting with an incorrect hypothesis, namely that one must manipulate both recessive and dominant genotypes resulted in incorrect evidence generation. Furthermore, this participant did not only manipulate the variables from her hypotheses, but she manipulated many more variables, which were irrelevant to the task at hand. This is an example of how performing scientific reasoning processes incorrectly can influence the resulting argumentation quality.

She wrote down the following answer to the task: '*Yes, the statement is true.*'. This answer is an example of low argumentation quality which received a score of 0. The participant stated an incorrect claim, she did not provide any evidence to support that claim, nor used scientific principles to support her claim (reasoning). Her epistemic network is displayed in Figure 9 on the left in red. As indicated by thick red lines in Figure 9, this participant engaged only in the processes of incorrect hypothesis generation (HypothesisGenerationI) and evidence generation (correct and incorrect) during the entire inquiry task. We can also see that there were no connections to other scientific reasoning processes, self-regulation processes, or generic cognitive strategies.

7. Discussion

In this study, we investigated how the processes of scientific reasoning and self-regulated learning and the use of generic cognitive strategies impact argumentation quality. To obtain the underlying reasoning and regulation processes, we collected fine-grained process data of these processes using screen captures and think aloud protocols. We used ENA to model the temporal associations between self-regulatory and scientific reasoning processes in students showing high and low argumentation quality.

Students with high argumentation quality performed different scientific reasoning processes more often together and correctly than students with low argumentation quality. Moreover, students who had high argumentation quality were monitoring more often during scientific reasoning than students with low argumentation quality. In particular, the stronger connections between scientific reasoning and self-regulation processes in the high argumentation quality group indicate that self-regulation processes were more important in that group. In contrast, fewer processes were co-occurring in the low argumentation quality group in general, and these were performed more often incorrectly than in the high argumentation quality group.

7.1 Theoretical and Methodological Contributions

Previous research has emphasized the importance of self-regulation for successful inquiry learning (see, Quintana et al., 2005; White et al., 2009). These studies have outlined that the lack of self-regulation (e.g., monitoring, reflection) can be harmful to students' inquiry learning and scientific reasoning. This study contributes to understanding the relationship between scientific reasoning and self-regulation in two ways. First, previous research has discussed the role of self-regulation during inquiry at a theory level (e.g., Andersen & Garcia-Milla, 2017), or has reported correlations between the quality of self-regulation skills and improvement of inquiry skills (e.g., Pedaste et al., 2012), or has focused on supporting self-regulation during inquiry (e.g., Manlove et al., 2007; White & Frederiksen, 1998). Our findings corroborate and extend previous research by providing a fine-grained investigation of the interaction between these processes in real-time during inquiry.

Second, this study builds upon previous research findings (e.g., White et al., 2009) by relating the conjoint use of scientific reasoning and self-regulation processes to argumentation quality. Using ENA, we analyzed and demonstrated the temporal associations between different scientific reasoning and self-regulation processes, which showed that students do not perform these activities in isolation, for example, students generate and evaluate evidence at the same time. Next, we statistically compared these patterns of association in students with high and low argumentation quality. For example, students with high argumentation quality were monitoring more often during evidence generation and evaluation in comparison with students with low argumentation quality. Last, we combined quantitative and qualitative data to demonstrate in which context self-regulation (e.g., monitoring) occurs spontaneously together with scientific reasoning processes and how it relates to students' argumentation.

Our findings corroborate previous research showing that the ability to self-regulate can be helpful for complex learning outcomes like argumentation and scientific reasoning. Engaging in metacognitive processes can improve both self-regulation abilities and scientific reasoning (White et al., 2009). Furthermore, using process data of students' scientific reasoning activities, we could demonstrate that monitoring during scientific reasoning activities was associated with high argumentation quality. Therefore, we could show that some students were engaging in self-regulation processes during scientific reasoning activities, which was related to high argumentation quality. Nevertheless, the frequency of spontaneous self-regulation was small, which points toward the importance of integrating instruction of self-regulation with instruction of scientific reasoning (White et al., 2009).

More frequent co-occurrences of different scientific reasoning processes were also found in the high argumentation quality group. These findings indicate that scientific reasoning processes did not occur in isolation from each other. Typically, scientific reasoning is seen as a cycle (e.g., inquiry cycle) and often research stipulates that this is a simplistic view of scientific reasoning, because scientific reasoning can begin at each step of the cycle (Hetmanek et al., 2018; White et al., 2009). What our findings demonstrate is that, instead of proceeding between different scientific reasoning steps sequentially during inquiry, students tend to perform different scientific reasoning activities conjointly in time. Therefore, we could reveal reciprocal links between different scientific reasoning processes, for example, evaluating evidence while drawing conclusions or generating evidence and evaluating it at the same time. Performing these activities simultaneously (and correctly) was observed more frequently in students with high argumentation quality.

Using advanced modeling methods allowed us to model how scientific reasoning and self-regulation processes impact argumentation quality. Using ENA proved useful for identifying emerging patterns of co-occurrence between different scientific reasoning processes and between self-regulation and scientific reasoning. Therefore, considering the temporal context of learning can reveal unique interactions between learning processes (Shaffer, 2017). Using ENA, we created quantitative models of qualitative data and went back to the specific think aloud transcripts to explain the meaning behind specific quantitative connections. Furthermore, we modeled the interaction between scientific reasoning, self-regulation, and generic cognitive strategies, such as self-explaining, and compared them in groups showing high and low argumentation quality.

ENA reveals specific participants which are responsible for the relationships we find. This allows us to go back to our original data and provide qualitative descriptions of the quantitative relationships we obtained. In this way, we can examine the reliability of our epistemic network models and inspect whether our findings are grounded in the original data (Shaffer, 2017). For example, ENA showed a stronger relationship between generating an incorrect hypothesis and correct self-explaining in the low argumentation quality group. Supplementing ENA findings with excerpts from the think aloud transcripts showed that correctly self-explaining *after* generating an incorrect hypothesis helped a student realize and correct their incorrect assumption. On the one hand, this might indicate that the correct self-explanations were tied to isolated facts that students knew, but they lack a comprehensive mental model of all relations. On the other hand, participants could have only verbalized information that they knew was correct but did not verbalize information they were unsure of. Both explanations are in line with the knowledge in pieces framework (diSessa, 1988, 2019), which postulates that deep knowledge consists of integrated pieces of knowledge in a mental model, but often students lack this integrated model. This could explain why despite providing correct self-explanations, participants did not integrate these pieces of knowledge in their final argumentation. The influence of self-explaining during inquiry learning requires further study.

ENA showed a strong co-occurrence of incorrectly and correctly generating evidence in the low argumentation quality group. This relationship indicates a violation of the control-of-variables strategy which also occurred more often in the low argumentation quality group. If we look only at the frequencies of correctly and incorrectly generating evidence, we would assume that the frequencies are unrelated. When we model their temporal association, we observe that these two processes were co-occurring together often and are resulting from control-of-variables strategy violations. Therefore, we argue that modeling the co-occurrences between processes provides a more comprehensive picture of the interaction between self-regulation and scientific reasoning in relation to argumentation quality.

7.2 Educational Implications

By taking a closer look at how university students approach inquiry learning using virtual experiments, we observed that students are still struggling with using the control-of-variables strategy. On one hand, this could result from a lack of scientific reasoning ability, indicating that students are unaware that by manipulating several variables at once, their results are confounded, and reliable conclusions cannot be drawn. On the other hand, virtual experiments on complex topics like population genetics provide many possible parameters which can be varied (see Figure 6a). This can result in students trying to "game the system" by manipulating multiple parameters to quickly obtain what they perceive to be the desired result, without trying to use the system for actual learning (Baker et al., 2008). Given another recent push to incorporate inquiry learning in science education, we argue that it is important to make students more aware of their actions during inquiry. We showed that when some students were self-explaining or monitoring, they realized their mistakes and could correct themselves. Asking students to self-explain their actions and reasoning during inquiry might alleviate the frequency of random manipulation of parameters, simply because the option is available. Self-explaining can potentially help students identify mismatches between their existing knowledge and their findings during inquiry and correct misconceptions in their mental models (deLeeuw & Chi, 2003). Learning gains from self-explanations depend on their quality. High-quality self-explanations that include generating inferences, integration, and deep processing (cf. principle-based explanations; Renkl, 1997) are associated with higher learning gains (Roy & Chi, 2005). Learners tend to struggle with providing high-quality explanations (Renkl, 1997; 2002) or they provide incorrect self-explanations, which can result in incorrect knowledge. In future research, the potential influence of self-explaining during inquiry should be investigated further.

Our ENA findings provide several educational implications for scientific reasoning and self-regulation instruction during inquiry learning. First, our findings showed that students were engaging in different scientific reasoning processes simultaneously. This implies that instruction on scientific reasoning should aim to foster the connection between different scientific reasoning activities (e.g., generating evidence while also evaluating evidence). Second, we provided insight into the importance of self-regulation during scientific reasoning activities for argumentation quality, which showed that students who were monitoring while performing scientific reasoning activities had better argumentation quality. Furthermore, not all students were self-regulating spontaneously, which indicates that they need self-regulation support during inquiry learning. Therefore, designing support measures that integrate instruction on self-regulation and scientific reasoning might prove to be beneficial for enhancing students' scientific reasoning and argumentation.

One effective method for enhancing scientific reasoning is providing instruction in the form of video modeling examples, which explain to students how to perform scientific reasoning activities correctly (Kant et al., 2017; Mulder et al., 2014). Furthermore, we argue that instruction on scientific reasoning should be integrated with instruction on self-regulated learning (planning, monitoring, reflecting). Video modeling instruction can be combined with narrative explanations, which demonstrate self-regulation during scientific reasoning. For example, metacognitive monitoring can be demonstrated by showing and correcting errors during inquiry. Furthermore, teaching students to plan their experiments at the beginning of the inquiry session could reduce "gaming the system" behaviors and help students design their experiments correctly. Instructing or prompting students to monitor their actions and their learning during the inquiry learning task could help them to become aware of performing scientific reasoning incorrectly and correcting their approach. Designing adaptive support tools, which remind students to connect different scientific reasoning processes, based on their actions in a virtual experiment, might be also beneficial for improving scientific reasoning. Future research needs to investigate the effects of different interventions (such as video modeling examples and self-regulation instruction) on improving scientific reasoning and argumentation.

7.3 Limitations

The main limitation of this study is the use of only one virtual experiment to study scientific reasoning, which limits the generalizability of the findings. We decided to use virtual experiments since they allow us to precisely trace and analyze fine-grained process data of scientific reasoning. We used a virtual experiment on the topic of genetics, which is representative of biology, and it allowed us to explore students reasoning on a topic that is a part of the biology curriculum. Nevertheless, a replication of these findings using a virtual experiment from a different topic in biology or a different domain (e.g., physics) would greatly improve the validity of our findings. Furthermore, studying scientific reasoning processes using virtual experiments allowed us to mainly investigate the processes of evidence generation and evidence evaluation, as indicated by higher frequencies of these processes in our data. Our task asked students to perform experiments to validate a hypothesis, which required them to mainly engage in evidence generation, evidence evaluation, and drawing conclusions. To further establish the validity of our findings, more research using tasks that evoke other scientific reasoning processes is needed. For example, asking students to identify the problem and generate a hypothesis by themselves before conducting experiments would allow for a more complete investigation of scientific reasoning processes.

Another limitation of this study was using a university sample to investigate the processes of self-regulation and scientific reasoning. However, we argue that it is important to investigate university students' scientific reasoning, as adults are still struggling with scientific reasoning (e.g., Smith & Wenk, 2006). Still, a replication of these findings with secondary school students would substantiate the validity of our findings. Furthermore, a comparison between the epistemic networks of university students and secondary school students would provide interesting additional evidence about the development of scientific reasoning. These are all questions that we are addressing in future research.

One limitation of our conceptualization of argumentation quality is that we did not consider rebuttals in our coding of argumentation. As a result, our definition of argumentation quality did not capture all dimensions of argumentation quality. Arguments consisting of multiple components are considered to be of higher quality (Osborne et al., 2004). Furthermore, there is no consensus in the field of argumentation on how to best define argumentation quality.

Another limitation of our study is the use of think aloud protocols to measure self-regulated learning. Previous research has shown that asking participants to think aloud might lead to reactivity (e.g., Chung et al., 2002), while the opposite was found by others (Bannert & Mengelkamp, 2008; Tarchi, 2021). Our findings that self-explaining helped some participants realize that their hypothesis was incorrect could indicate potential reactivity caused by the think aloud method, or it could be the result of self-explaining itself. More research using different methods of assessing self-regulation (e.g., behavioral measures) in relation to scientific reasoning is needed.

Last, there were several issues with the internal consistency of the scientific reasoning ability and prerequisite conceptual knowledge tests that we used. Consequently, we were not able to reliably assess scientific reasoning ability and prerequisite conceptual knowledge. However, we argue that looking at scientific reasoning processes might be an informative alternative way of measuring students' actual scientific reasoning abilities (Gobert et al., 2012). Participants in the low argumentation quality had slightly higher domain-specific interest, which was unexpected given the positive influence of interest on learning processes typically found. However, due to the small sample size and the relatively high p-value, more research is needed to replicate and explain this finding. Future research should investigate how actual inquiry behavior is related to (reliable) scientific reasoning ability tests and how prerequisite conceptual knowledge is associated with scientific reasoning processes. When investigating the role of prerequisite knowledge in this context, it is important to also consider the influence of the different aspects of prior knowledge, namely its accuracy, amount, specificity, and coherence (McCarthy & McNamara, 2021).

8. Conclusion

The present study provided a first attempt to investigate how scientific reasoning and self-regulatory processes impact argumentation quality. Our findings indicated that students who showed high argumentation quality performed different scientific reasoning processes better and they also made more connections between self-regulation and scientific reasoning. Furthermore, we observed rare instances of spontaneous self-regulation processes. Monitoring during scientific reasoning activities was more often present in participants who had high argumentation quality. These findings offer educational implications for teaching scientific reasoning. Instruction which integrates self-regulation and scientific reasoning would be beneficial for improving inquiry learning.

VIII. Study 3: Do Video Modeling and Metacognitive Prompts Improve Self-Regulated Scientific Inquiry?

1. 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.

Published as: Omarchevska, Y., Lachner, A., Richter, J., & Scheiter, K. (2022). Do video modeling examples and metacognitive prompts improve self-regulated scientific inquiry? *Educational Psychology Review,* https://doi.org/10.1007/s10648-021-09652-3

2. Introduction

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.

3. Theoretical Framework

3.1 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.

3.2 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).

3.2.1 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 & 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.

3.2.2 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, 2009b). 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.

3.2.3 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, b) 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, b). 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, b), 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.

3.2.4 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).

3.3 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.

3.3.1 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.

3.3.2 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.

4. 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.

5. Method

5.1 Participants and Design

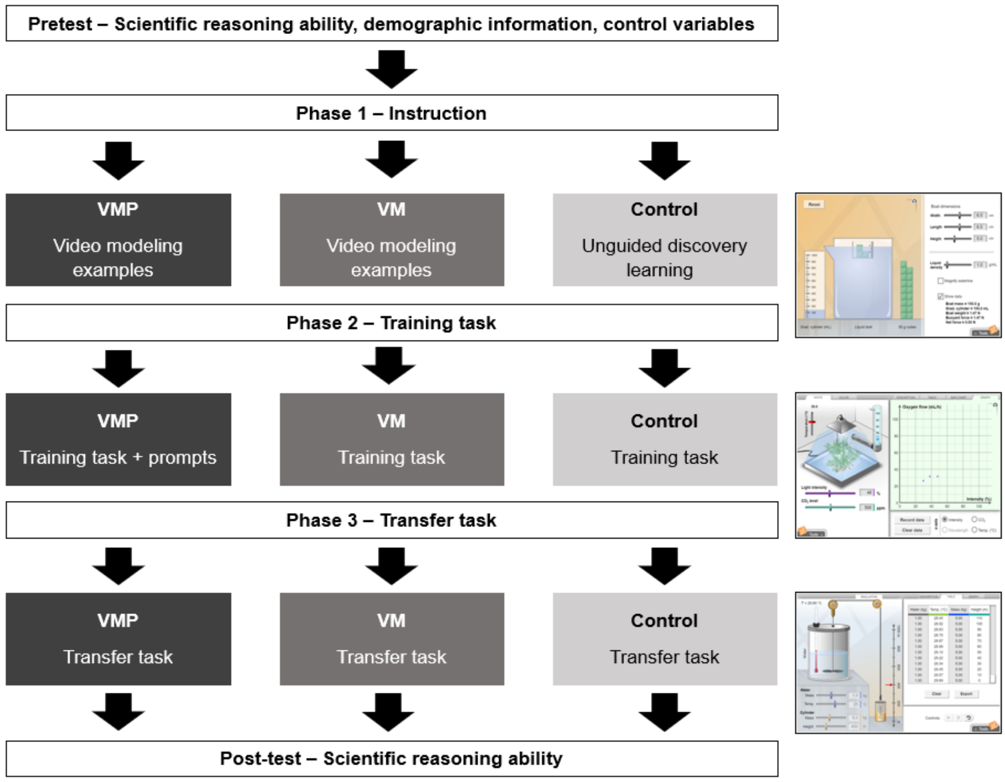
Participants were 127 university students from Southern Germany (26 males, M_{age} = 24.3 years, *SD* = 4.81). Participants had an academic background in science (*n* = 40), humanities (*n* = 43), law (*n* = 11), social science (*n* = 20), or other (*n* = 14). Participation in the experiment was voluntary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

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 Figure 10). 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 = .80, α = .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.

Figure 10

The Design and Procedure of the Experiment Divided in Three Phases and the Corresponding Virtual Experiment Used in Each Phase



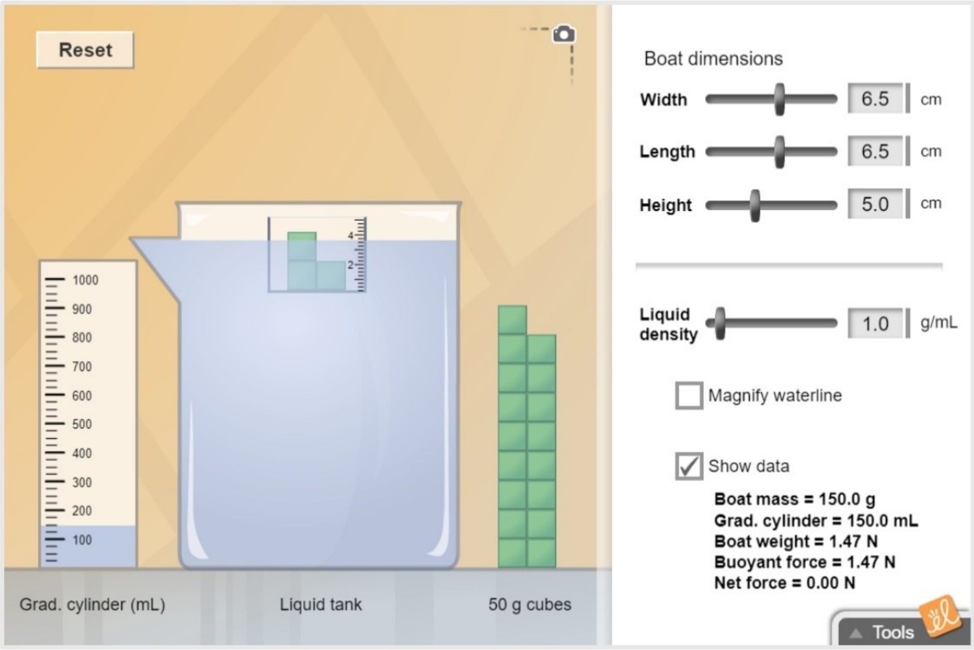
5.2 Materials and Procedure

5.2.1 Phase 1 – Instruction. We first assessed demographic information, conceptual knowledge, academic interest and self-concept (Figure 10). 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, Figure 11). 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 minutes 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.

Figure 11

The Archimedes' Principle Simulation



Note. Copyright (2021) by Explore Learning. Reprinted with Permission.

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 minutes to work on the task.

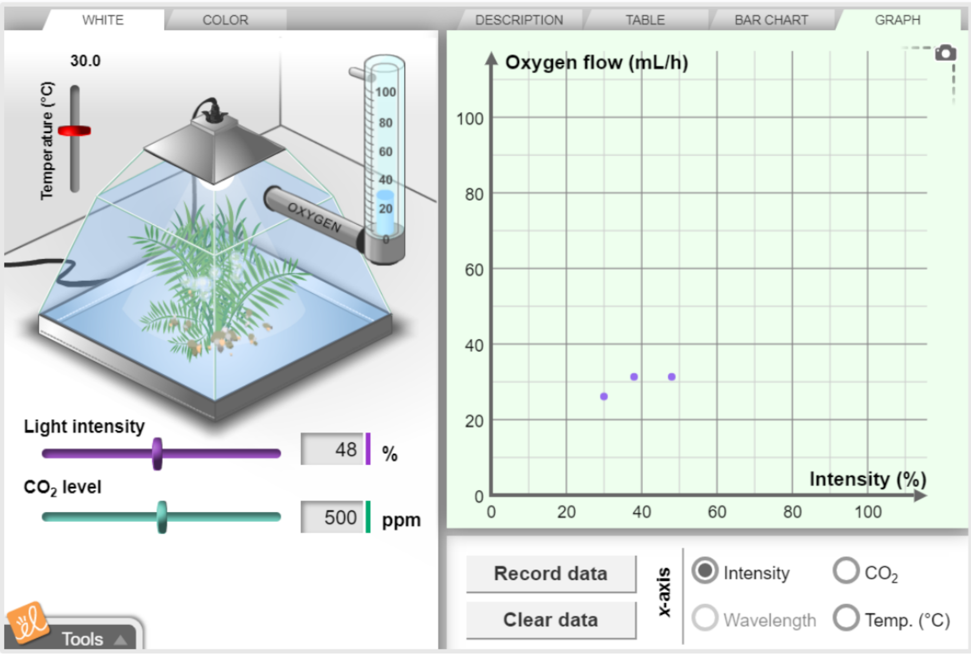
5.2.2 Phase 2 – Training Task. Participants were first instructed to think aloud by asking them to say everything that comes to their mind without worrying about the formulation. Participants were given a short practice task ("Your sister's husband's son is your children's cousin. How is he related to 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 Figure 12), 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 8), which asked them to 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 minutes, respectively. The third prompt was visible after participants finished the training task and gave the option to go back and conduct more experiments.

Figure 12

The Photosynthesis Simulation



Note. Copyright (2021) by Explore Learning. Reprinted with Permission.

Table 8

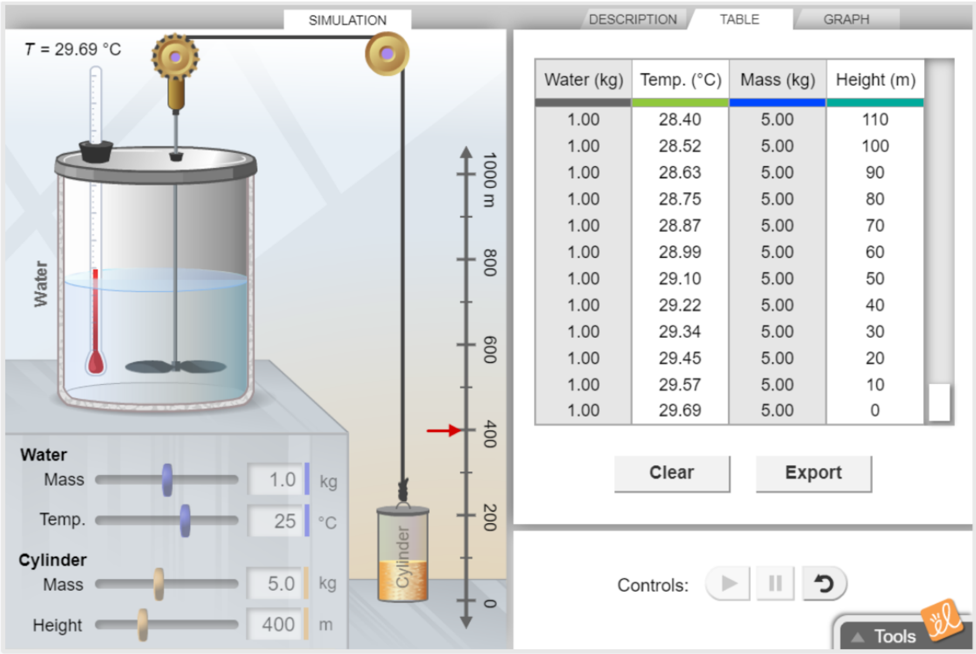
The Metacognitive Prompts Used in the Video Modeling and Prompting Condition.

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?

5.2.3 Phase 3 – Transfer Task. In the transfer task, all participants worked with the *Energy Conversion in a System* (see Figure 13) 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 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.

Figure 13

The Energy Conversion in a System Simulation



Note. Copyright (2021) by Explore Learning. Reprinted with Permission.

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.

5.3 Measures

5.3.1 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 α = .63; energy conversion, Cronbach's α = .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).

5.3.2 Academic Self-Concept and Interest in Science. The academic self-concept scale comprised 5 items rated on a Likert scale (Cronbach's α = .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 α = .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.".

5.3.3 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 α = .31), most likely because the test assessed three independent skills (cf. Stadler et al., 2021).

5.3.4 Scientific Reasoning Products.

5.3.4.1 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 9. 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 α = .65). Two raters independently scored all hypotheses ($N = 127$) and then discussed all disagreements until a consensus was reached.

5.3.4.2 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 10. 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 $\alpha = .67$) for the scoring.

5.3.5 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 11, 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*.

5.3.6 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 0-100. 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).

5.4 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 $\alpha = .025$ for all contrast analyses. Benchmarks for effect sizes were: $\eta^2 = .01, .06, .14$ 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.

5.4.1 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.

Table 9

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.	The hypothesis is formulated in a way that can be tested. Dependent and independent variables are specified. If-then statements are used, and the relationship is specified.
Example from data	*Under what conditions does the plant produce its maximum amount of oxygen?*	*Light intensity influences oxygen production during photosynthesis.*	*When I increase the light intensity then oxygen production will increase as well.*
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).	The hypothesis is correct.
Example from data	*Plants need light for photosynthesis.*	*Temperature has an effect on oxygen production*. (irrelevant variable)	*The higher the light intensity, the higher the oxygen production during photosynthesis.*

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

Table 10

Coding Scheme Used for Assessing Argumentation Quality

	Level		
Component	*0*	*1*	*2*
Claim			
Criteria	Does not make a claim or makes an inaccurate claim.	Makes an accurate but incomplete claim. Includes irrelevant variables.	Makes an accurate and complete claim.
Example from data	*The example shows that a lot of oxygen was produced at full light intensity (100%).* (inaccurate claim)	*Depending on the temperature, the curve of oxygen production runs longer or shorter linearly before it stagnates.* (irrelevant variable: temperature)	*If the light intensity is increased above 40%, the oxygen production remains unchanged at 30 mL/h.*
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)	*oxygen production increases when certain temperature (perfect 24-28), right light intensity (50-60%), quite high CO_2 level (from 420 ppm)* (appropriate and inappropriate evidence)	*The light intensity influences the oxygen flow linearly up to a value of 37% under constant conditions.*
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	*The initial temperature has an influence on the temperature at the end of the water and on the mass.*	*With a light intensity of 60% we have a maximum O_2 production.*	*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.*

Table 11

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.

5.4.2 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 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.

6. Results

6.1 Control Variables

A MANOVA with the control variables (conceptual knowledge, interest, academic self-concept) revealed no differences between conditions, $F < 1$, see Table 12 for descriptive statistics.

6.2 Scientific Reasoning Ability

A contrast analysis with scientific reasoning ability (pretest) as a covariate revealed no group differences in the posttest scores (Contrast 1, β = 0.04, p = .57, d = 0.22; Contrast 2, β = -0.02, p = .76, d = 0. Thus, there was no support for H1a and H1b[3].

Table 12

Descriptive Statistics (and SDs) for the Control and Dependent Variables, and Monitoring Accuracy (Bias)

Variable	VMP (n = 43)	VM (n = 43)	Control (n = 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.

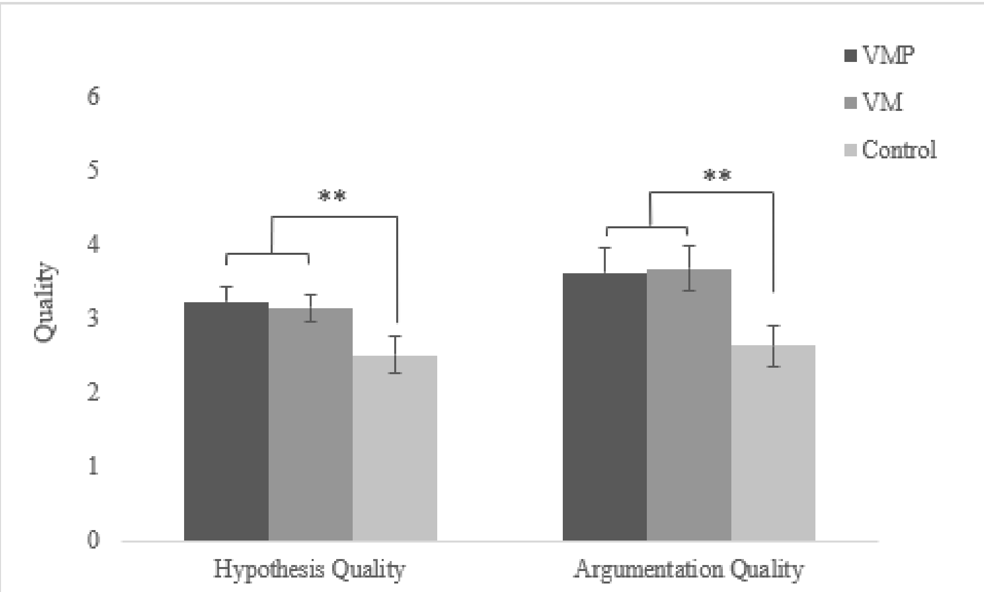
[3] Additional analyses revealed that time on task was significantly shorter for the posttest, $F(1, 121) = 475.21$, $p < .001$, Wilk's $\Lambda = .203$, $\eta_p^2 = .79$, suggesting that participants clicked through the posttest very quickly. Therefore, this finding should be interpreted with caution.

6.3 Effects on the Training Task

6.3.1 Product Level. Contrast 1 showed that the VMP and VM conditions had higher hypothesis quality, $t(124) = 2.60$, $p = .010$, $d = 0.49$, and higher argumentation quality, $t(124) = 2.75$, $p = .007$, $d = 0.52$, than the control group, see Figure 14. Contrast 2 showed no significant differences between the VMP and VM in hypothesis quality, $p = .75$, $d = 0.02$, and argumentation quality, $p = .91$, $d = 0.02$. 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, $p = .20$, $\eta^2 = .03$, and argumentation quality, $p = .20$, $\eta^2 = .03$.

Figure 14

The Immediate Training Effects of Video Modeling (VM) and Metacognitive Prompts (P) on Hypothesis and Argumentation Quality



Note. Error bars show standard errors.

6.3.2 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 13.

6.3.2.1 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* = .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* = .14, *d* = 0.37).

Thicker green lines in Figure 15 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.

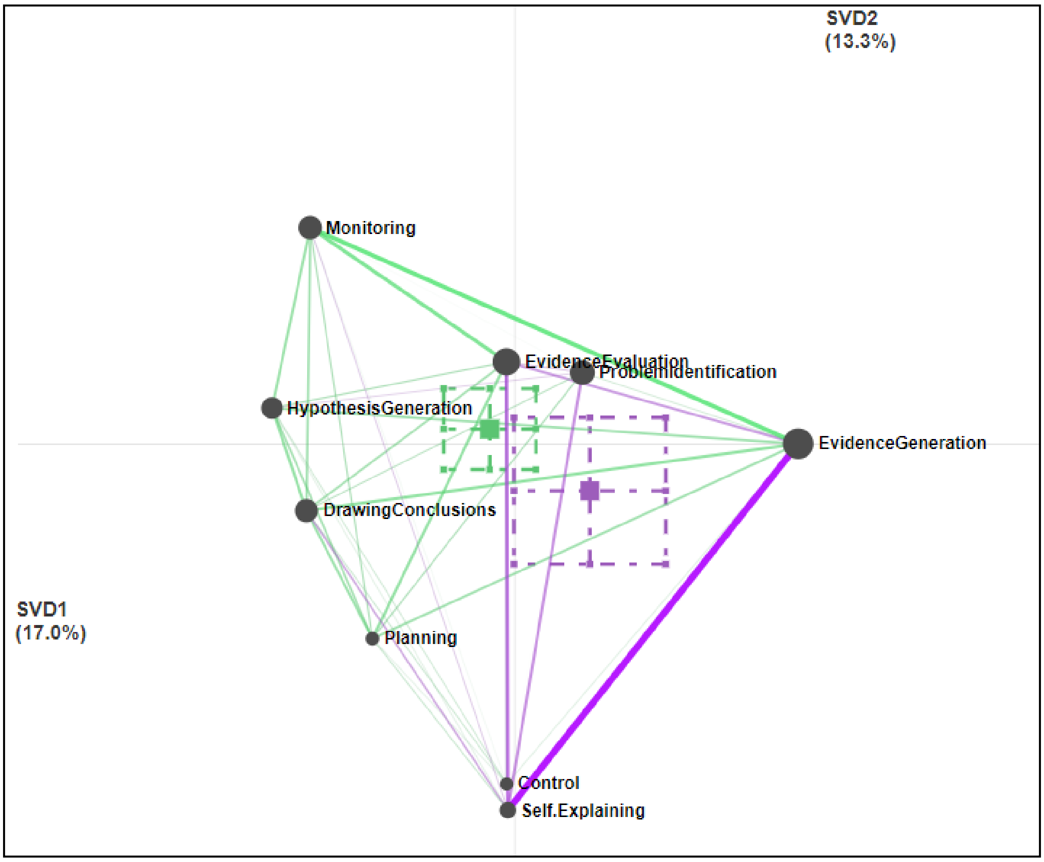
Table 13

The Total Frequencies (and Percentages) of Process Events per Condition in the Training Task

	VMP (*n* = 29)	VM (*n* = 37)	Control (*n* = 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

Figure 15

The Epistemic Network Difference Between the VMP and VM (left, green) Conditions and the Control (right, purple) Condition in the Training Task

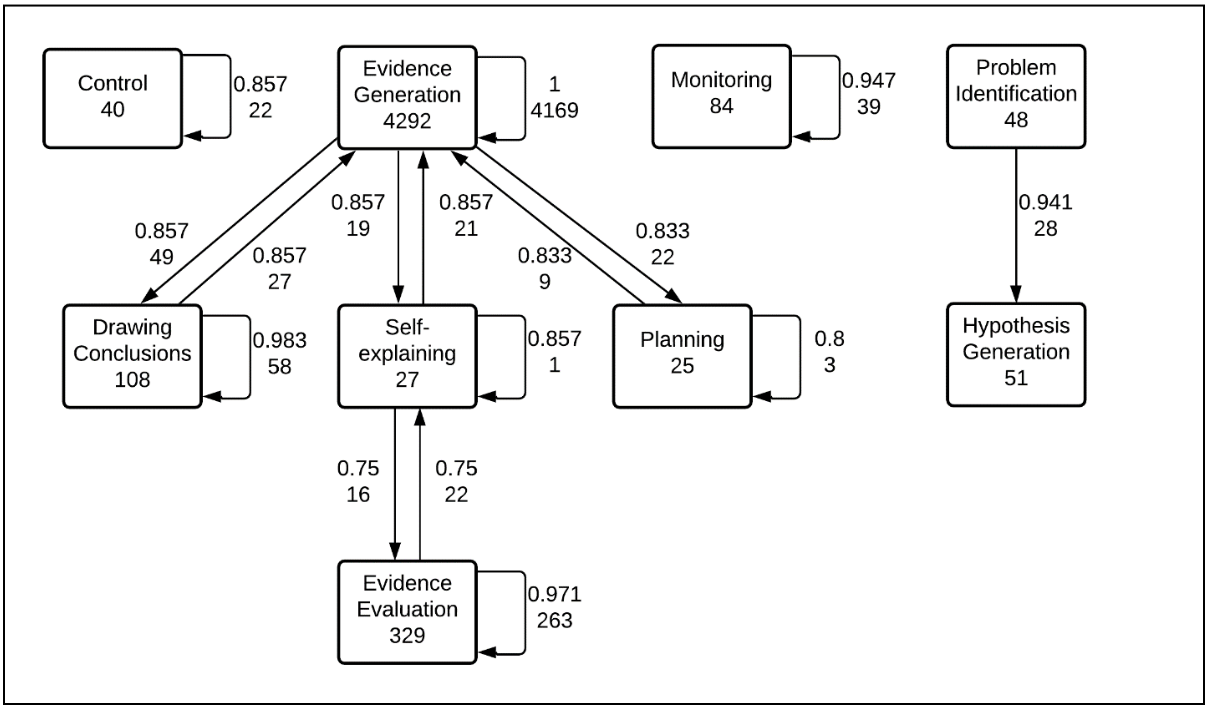


6.3.2.2 ENA: VMP vs. VM. Two sample-*t* tests revealed no differences between the position of the VMP centroid compared to the VM control along either the X axis, $t(55.53) = 0.52$, $p = .60$, $d = 0.13$) or the Y axis, $t(59.54) = -0.42$, $p = .67$, $d = 0.11$. Thus, the epistemic networks of the groups with video modeling were similar (for details see Appendix B).

6.3.2.3 Process Mining. The process model of the two video modeling conditions in the training task (Figure 16) 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.

Figure 16

The Process Model of the VMP and VM Conditions Combined (n = 66) in the Training Task

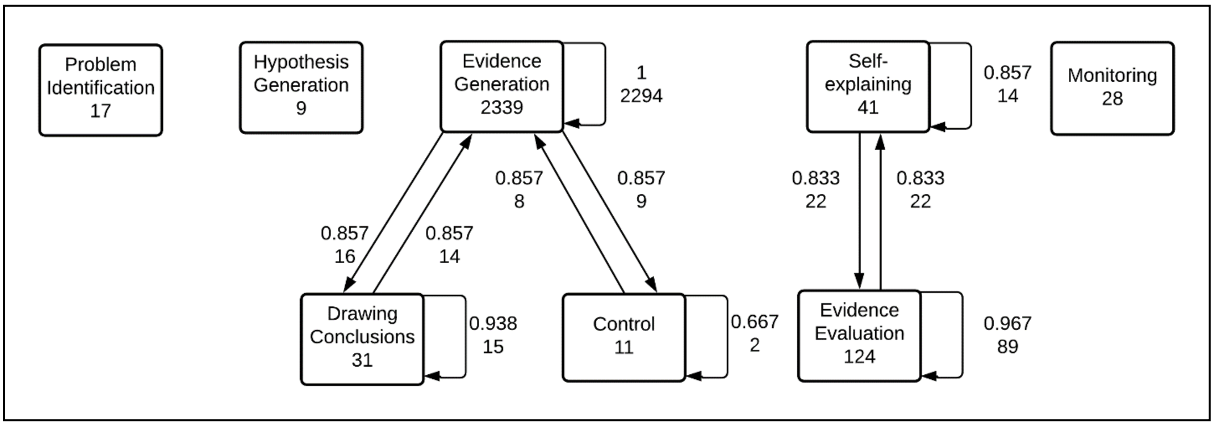


Note. The top number (referring to the arrows) represents the dependency (0-1), and the number below represents the frequency of each sequence. The numbers in the boxes represent the frequency of each process.

In contrast, in the control condition (Figure 17), 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.

Figure 17

The Process Model of the Control Condition (n = 22) in the Training Task



Note. 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.

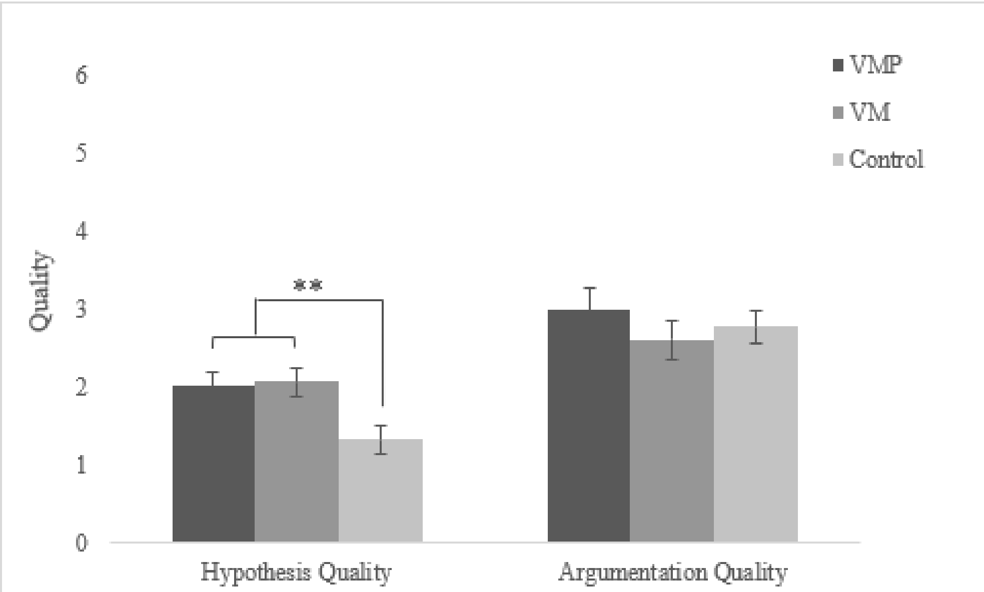
6.4 Effects on the Transfer Task

6.4.1 Product Level. Contrast 1 indicated that VMP and VM had higher hypothesis quality than the control condition, $t(124) = 3.06$, $p = .003$, $d = 0.58$, but they did not differ significantly from each other (Contrast 2, $t(124) = -0.22$, $p = .83$, $d = 0.05$), see Figure 18. The three groups did not differ in argumentation quality (Contrast 1, $t(124) = 0.09$, $p = .93$, $d = 0.02$; Contrast 2, $t(124) = 1.09$, $p = .28$, $d = 0.22$). 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, $F(2, 122) = 5.68$, $p = .004$, $\eta^2 = .09$. Post-hoc comparisons with Bonferroni corrections indicated that participants in the control condition overestimated their hypothesis quality, compared to the VMP, $p = .006$, and the VM condition, $p = .028$. There were no differences in monitoring accuracy for argumentation quality, $p = .20$, $\eta^2 = .02$.

Figure 18

The Transfer Effects of Video Modeling (VM) and Metacognitive Prompts (P) on Hypothesis and Argumentation Quality



Note. Error bars show standard errors.

6.4.2 Process Level. Descriptive statistics are reported in Table 14.

Table 14

The Total Frequencies (and Percentages) of Process Events per Condition in the Transfer Task

	VMP (*n* = 29)	VM (*n* = 37)	Control (*n* = 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

6.4.2.1 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 control along either the X axis, $t(34.29) = -0.63$, $p = .53$, $d = 0.16$, or the Y axis, $t(34.90) = -1.43$, $p = .16$, $d = 0.36$. Thus, the epistemic networks of the groups with and without video modeling were similar (for details see Appendix B).

6.4.2.2 ENA: VMP vs. VM. Two sample *t*-tests showed no significant differences between the centroids of the VMP and VM conditions along the X, $t(57.37) = -1.62$, $p = .11$, $d = 0.40$, and the Y axis, $t(54.76) = -0.43$, $p = .67$, $d = 0.11$. 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 Appendix B).

7. 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 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.

7.1 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., Manlove et al., 2009b; 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, b) 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, b) 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.

7.2 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).

7.3 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 minutes), 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.

7.4 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.

8. 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.

IX. Results and Discussion

Educational reforms in the past decades (e.g., KMK, 2020; NASEM, 2021; NRC, 1996, 2013) have pushed toward the increased implementation of active learner-centered methods like inquiry learning in science education to improve students' scientific reasoning and argumentation. Because the effectiveness of active learner-centered approaches relies on students' active engagement and participation (Freeman et al., 2014), there are important factors that need to be considered for their effective implementation. The three conceptual aims of this dissertation related to investigating 1) the conjoint role of students' cognitive and motivational characteristics on their experimentation skills and conceptual understanding, 2) the interplay of self-regulation and scientific reasoning processes and argumentation quality, and 3) the effectiveness of integrated scaffolding of self-regulation and scientific reasoning for students' learning processes and learning outcomes.

In Study 1, a person-oriented approach was applied to investigate the combined role of students' cognitive and motivational prerequisites for their experimentation skills and conceptual understanding. Findings showed that different configurations of motivational and cognitive prerequisites need to be considered during inquiry learning. In Studies 2 and 3, data on the conjoint and sequential use of scientific reasoning and self-regulation processes were collected, in addition to learning outcomes, to investigate their interplay. Results from Study 2 indicated that self-regulation processes, in addition to scientific reasoning processes, need to be considered to better understand inquiry learning. Study 3 reported the effects of an intervention, designed based on the findings from Study 2, targeted at supporting the integrated use of self-regulation and scientific reasoning. In Study 3, prior to engaging in inquiry, instruction on the integrated use of self-regulation and scientific reasoning was provided using video modeling examples. Self-regulation during scientific reasoning was additionally supported using metacognitive prompts. Findings from Study 3 showed the effectiveness of the integrated instruction of scientific reasoning and self-regulation. An additional methodological contribution of this dissertation is providing insight into the role of self-regulated learning during scientific reasoning tasks using advanced methods which capture the conjoint and sequential use of these processes.

In the following, the results of the three studies are summarized. Next, the theoretical implications of the studies are discussed. To this end, their findings are integrated into a theoretical model (Self-Regulated Learning during Scientific Reasoning [SRLSL] Model), which illustrates the relationships between students' individual prerequisites, self-regulation, scientific reasoning processes, and learning outcomes. Last, the educational implications, strengths, and limitations of the studies are addressed, and an overall conclusion is drawn.

1. Summary of Results

To address open questions regarding the role of students' characteristics, self-regulation, and integrated guidance of self-regulation and scientific reasoning in the context of inquiry learning, three studies were conducted which are summarized in the following section.

Study 1 investigated how cognitive and motivational characteristics conjointly relate to students' experimentation skills and conceptual understanding in a guided inquiry learning science lesson. Secondary school students from six classes (N = 110, M_{age} = 12.07 years) attended a lesson on photosynthesis and solved a structured inquiry task using a simulation about photosynthesis individually on a laptop. Experimentation skills were measured by the proportion of controlled experiments obtained from screen recordings of students' inquiry behavior. At the end of the task, students wrote an explanation answering the research question they investigated, from which argumentation quality was coded. Conceptual understanding comprised students' written argumentation quality and posttest scores. Results from a *k*-means cluster analysis revealed three distinct clusters, based on students' prior knowledge, academic self-concept, and interest – Underestimating, Struggling, and Strong. Underestimating students had high prior knowledge, but low interest and self-concept. Struggling students scored below average on all three variables, whereas Strong students scored above average on all variables. Regarding experimentation behavior, surprisingly, Struggling students performed a significantly higher proportion of controlled experiments than Strong students. Underestimating and Strong students scored significantly higher than Struggling students in the conceptual knowledge posttest. The three clusters did not differ significantly in argumentation quality. Last, a mediation analysis indicated that experimentation skills did not mediate the relationship between students' individual prerequisites and conceptual understanding, suggesting that the proportion of controlled experiments might be a too limited measure of experimentation skills.

To further investigate factors related to students' experimentation, Study 2 addressed open questions arising from Study 1. First, to investigate scientific reasoning processes going beyond experimentation skills, additional epistemic activities that students engaged in were coded and analyzed. Second, self-regulation processes, obtained from think aloud protocols, were investigated. Third, to investigate how the co-occurrences of self-regulation and scientific reasoning processes relate to argumentation quality, epistemic network analysis (ENA) was used to model the conjoint use of these processes. To observe how self-regulation and scientific reasoning processes occur without scaffolds, only minimal guidance was provided. University students ($N = 30$, $M_{age} = 23.33$ years) solved an inquiry task that asked them to test an assumption made by a fictitious person using a computer simulation on the topic of population genetics. At the end of the task, they wrote an explanation regarding the original research question, from which argumentation quality was coded. Fine-grained process data, obtained from screen recordings and think aloud protocols during the inquiry task, were used to explore the co-occurrences of scientific reasoning and self-regulation processes in relation to argumentation quality. A comparison between students with high and low argumentation quality using ENA revealed important differences between the two groups on the process level. First, students with high argumentation quality performed different epistemic activities conjointly and correctly more frequently than students with low argumentation quality. For instance, high argumentation quality was associated with generating and evaluating evidence in the same temporal context, whereas low argumentation quality was associated with violations of the control of variables strategy. Moreover, high argumentation quality was associated with more frequent co-occurrences of self-regulation processes and epistemic activities. In particular, participants with high argumentation quality in their answers were monitoring while generating evidence. These findings suggest that instruction that integrates self-regulation and self-regulation could be beneficial for improving scientific reasoning and argumentation.

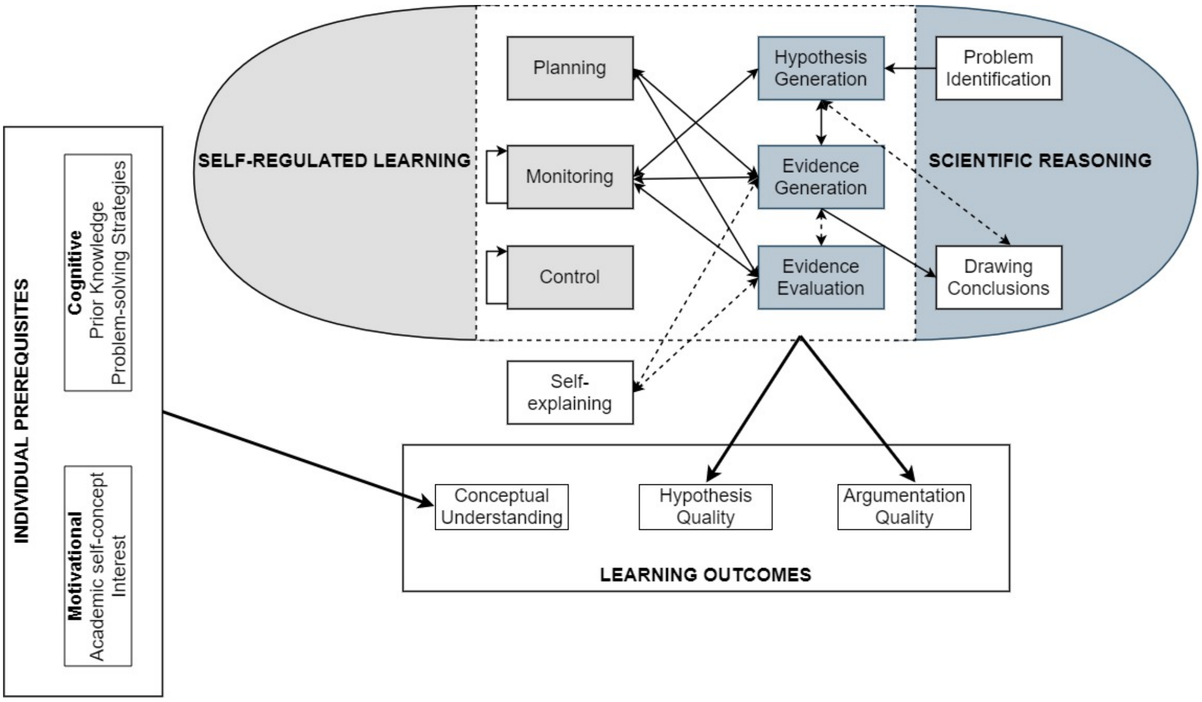
Accordingly, Study 3 investigated the effectiveness of video modeling examples (VM) and metacognitive prompts (P) as an integrated instruction of self-regulation and scientific reasoning. The video modeling examples and prompts were designed based on the findings from Study 2 and aimed to foster the self-regulation of specific epistemic activities (e.g., evidence generation, evidence evaluation). The effects of watching VM examples combined with prompts (VMP) to watching VM examples only, and to unguided inquiry (control) were compared in two inquiry tasks (training and transfer task). The effectiveness of the intervention was evaluated regarding university students' ($N = 127$, $M_{age} = 24.3$ years) scientific reasoning ability, hypothesis and argumentation quality, and scientific reasoning and self-regulation processes. Results indicated that 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 were no differences between the two video modeling conditions, suggesting 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. Using ENA and process mining, the co-occurrences and sequences of scientific reasoning and self-regulation were modeled. ENA identified stronger co-occurrences between scientific reasoning and self-regulation processes in the two VM conditions compared to the control condition. Specifically, monitoring and planning were used conjointly with hypothesis generation, evidence generation, and evidence evaluation. Process mining additionally revealed specific sequences of these processes. These findings indicate that an instruction that aimed at integrating self-regulation and scientific reasoning, resulted in more conjoint use of scientific reasoning and self-regulation processes, and higher hypothesis and argumentation quality. An integration of findings on the process and learning outcomes level revealed that the processes modeled in the video modeling examples (correct hypothesis generation, planning and monitoring scientific reasoning activities) resulted in improved hypothesis and argumentation quality and a conjoint use of self-regulation and scientific reasoning processes.

2. Theoretical Implications

The three conceptual goals of this dissertation refer to the role of students' individual prerequisites, self-regulated learning processes, and an integrated instruction of scientific reasoning and self-regulation on learning processes and outcomes during inquiry learning. The employment of advanced methods, which can capture the co-occurrence and sequential relationships between learning processes, allowed for an in-depth investigation of the interaction between self-regulation and scientific reasoning in relation to hypothesis and argumentation quality. The findings from the three studies are integrated into the Self-Regulated Learning during Scientific Reasoning (SRLSR) model, presented in Figure 19. The three major theoretical implications stemming from the theoretical model are discussed in the following section.

Figure 19

Model of Self-Regulated Learning during Scientific Reasoning (SRLSR Model)



The first implication of the research reported in this dissertation that is captured in the SRLSR model refers to the role of students' cognitive (prior knowledge) and motivational (academic self-concept and interest in science) prerequisites for learning outcomes, which was investigated in Study 1 (see left part of the model). Findings from Study 1 indicated that both cognitive and motivational variables were associated with conceptual knowledge, but not with argumentation quality. Despite the influence of motivation on students' engagement (Hidi, 1990), previous research on inquiry learning has largely neglected the role of prior motivational variables on students' engagement during inquiry learning. When investigating the relationship between inquiry learning and motivation, the reverse relationship is often considered in prior research, namely, the influence of engaging in inquiry learning activities on students' motivation and situational interest (e.g., Krüger et al., 2022; Rutten et al., 2012). According to findings from Study 1, both cognitive and motivational prerequisites are important for inquiry learning. Likewise, academic self-concept can also affect students' engagement (Schnitzler et al., 2021) and achievement (Huang, 2011). However, these findings are not from the specific context of inquiry learning. Similar to interest, prior research has investigated the influence of engaging in inquiry activities on students' academic self-concept (Jansen et al., 2015), but the reverse relationship has not been investigated. As a consequence, students who are motivated and perceive themselves as capable in science (high academic self-concept in science) might also benefit from different types of guidance than students who are unmotivated and perceive themselves as less capable in science (Huber & Seidel, 2018). So far, previous research has considered the influence of different levels of prior knowledge on the effectiveness of different types of guidance (van Riesen et al., 2018, 2019). However, findings from Study 1 suggest that it might also be effective to tailor guidance during inquiry learning to students' prior motivational characteristics, not only to their prior knowledge.

Findings from Study 1 also demonstrated that only considering the control-of-variables strategy (CVS) might be a too limited measure of experimentation skills and instead, additional cognitive and metacognitive processes need to be considered. Furthermore, relying only on behavioral data, coded from screen recordings, does not provide insight into the intentionality of students' experimentation behavior. Study 1 findings indicated that students with the lowest prior knowledge, self-concept, and interest (Struggling cluster) performed the highest proportion of controlled experiments which in contrast with prior research showing that low prior knowledge students tend to violate the CVS more frequently than high prior knowledge students (Bumbacher et al., 2018). However, the highly structured design of the inquiry task might have been beneficial for students with low prior knowledge, self-concept, and interest, but might not have been suitable for students with higher prior knowledge, self-concept, and interest (Kalyuga, 2007). These findings suggest that the way inquiry learning is implemented (task structuring can be seen as a form of guidance that supports students by restricting the number of variables to consider for their experiments) needs to be tailored to students' motivation, in addition to their prior knowledge. Furthermore, students' cognitive and metacognitive processes need to be considered, to provide insight into the intentionality of their experimentation behavior, which cannot be obtained from simply looking at their inquiry behavior. Therefore, Studies 2 and 3 focused on students' scientific reasoning processes (i.e., epistemic activities extending beyond the control-of-variables strategy) and their self-regulation processes.

Second, the SRLSR model captures the evidence of the interplay between different scientific reasoning processes (i.e., epistemic activities) in relation to hypothesis and argumentation quality that was obtained in this dissertation. The conjoint and sequential use of different epistemic activities is illustrated on the top right of the SRLSR model by integrating findings from Studies 2 and 3. Solid lines indicate that these relationships among scientific reasoning and self-regulation processes were found after the interplay of self-regulated learning and scientific reasoning had been instructed using video modeling examples or were found in the high argumentation quality group in Study 2. Dashed lines illustrate relationships which were stronger either in the low argumentation quality group in Study 2 or the control condition of Study 3, which also had lower hypothesis and argumentation quality. Bi-directional arrows indicate a conjoint use of the processes, obtained from ENA findings. Unidirectional arrows illustrate that a specific sequence was identified by process mining, indicating that there is a higher likelihood of one process preceding the other.

In the following, the findings related to hypothesis and argumentation quality are discussed (solid lines illustrate the relationships regarding higher hypothesis and argumentation quality). The model illustrates that problem identification was followed by hypothesis generation, which is in line with inquiry learning and scientific reasoning frameworks (Fischer et al., 2014; Pedaste et al., 2015). Moreover, hypothesis generation and evidence generation co-occurred together, indicating that the two processes were performed together more frequently and were related to higher hypothesis and argumentation quality. The fact that problem identification was only connected to hypothesis generation has two potential explanations. First, in the design of the inquiry tasks, the problem was already identified, so participants started the inquiry process by generating a hypothesis. Second, problem identification is often the starting point of the inquiry process (Pedaste et al., 2015) and therefore is not performed later in the task after the problem has already been identified. Evidence generation was more frequently followed by drawing conclusions, which is also in line with frameworks of scientific reasoning and inquiry (Fischer et al., 2014; Pedaste et al., 2015). Dashed lines – which represent relations found in students with lower argumentation quality – indicate stronger co-occurrence between hypothesis generation and drawing conclusions. This finding might suggest that these students were generating hypotheses and drawing conclusions without engaging in core scientific reasoning processes (evidence generation and evidence evaluation) necessary for hypothesis testing and drawing conclusions.

The relationship between evidence generation and evidence evaluation is illustrated by a dashed line, indicating that these processes were performed more frequently conjointly in the control condition in Study 3, which, however, showed a rather weak relationship between the two processes. At the same time, a strong co-occurrence of evidence generation and evidence evaluation was found in the high argumentation quality group in Study 2. These mixed findings could be explained using the process mining graph of the experimental conditions in Study 3 (see Figure 16), which indicated a pattern of evidence generation, followed by self-explaining, followed by evidence evaluation. Therefore, a potential explanation of the weaker co-occurrence between evidence generation and evidence evaluation in the experimental conditions found using ENA is that, instead of simultaneously generating and evaluating evidence, these participants were also self-explaining in-between. ENA only considers the relationship between two processes, whereas process mining can identify longer sequences of processes.

Overall, findings indicate that performing more scientific reasoning processes conjointly is associated with higher argumentation quality. Previous studies have looked at the conjoint use of scientific reasoning processes using ENA (e.g., Bauer et al., 2020; Csanadi et al., 2018). One of the novel contributions of this dissertation is to relate the conjoint use of scientific reasoning processes to hypothesis and argumentation quality. That is, the higher number of connections between different scientific reasoning processes was found in the network of the two experimental conditions in Study 3, which also showed better inquiry performance, and in the higher argumentation quality network in Study 2.

The third theoretical implication captured in the SRLSR model relates to the in-depth investigation of the interplay of scientific reasoning and self-regulation processes in relation to hypothesis and argumentation quality. Prior research has emphasized the importance of self-regulation during inquiry learning and scientific reasoning (e.g., Manlove et al., 2009b; White et al., 2009). In prior research, either scientific reasoning (e.g., Liu et al., 2022) processes or self-regulation processes were coded (e.g., Manlove et al., 2007). However, how these processes interact during the inquiry tasks has not been previously investigated. Therefore, despite providing theoretical claims about the importance of self-regulation for scientific reasoning, little direct evidence exists about the interplay of these constructs. Findings from this dissertation address this research gap by providing direct evidence for the role of specific self-regulation processes for scientific reasoning during inquiry learning.

The findings that are represented in the SRLSR model indicate that monitoring was co-occurring with hypothesis generation, evidence generation, and evidence evaluation. Planning co-occurred with evidence generation and evidence evaluation. These relationships are illustrated using solid bi-directional lines, indicating that the conjoint use of these processes was observed in the two video modeling conditions in Study 3. Furthermore, it demonstrates that the integration of self-regulation and scientific reasoning processes, which was instructed using the video modeling examples in Study 3, resulted in a conjoint use of these processes. The conjoint use of self-regulation and scientific reasoning processes was also related to improved hypothesis and argumentation quality. Therefore, initial exploratory findings from Study 2, which showed that monitoring during scientific reasoning activities was observed more frequently in students with high argumentation quality, were confirmed experimentally on the process and learning outcomes level.

Next to metacognitive processes, the use of other cognitive strategies, like activation of prior knowledge and self-explaining, were considered in Studies 2 and 3. Overall, activation of prior knowledge occurred very rarely and thus was excluded from the process analyses and the SRLSR model. Self-explaining co-occurred more frequently with evidence generation and evidence evaluation in the control condition of Study 3, indicating that students who did not engage in self-regulation processes relied on self-explaining as an additional learning strategy during evidence generation and evidence evaluation, which was associated with lower argumentation quality. In contrast, findings from Study 2 showed that self-explaining can also be beneficial for students, as some students realized that their initial assumptions were incorrect and could correct themselves. Nevertheless, for self-explaining to be beneficial, the explanations should be correct, and their quality should be high (Renkl, 1997). The role of self-explaining should be further investigated, for example, by instructing students to provide self-explanations that include deep processing and integration of information, which could then possibly benefit students struggling with self-regulation.

So far, the SRLSR model does not include relationships between control and any other processes, as there was no evidence for these relations in the present studies. One potential explanation for this finding is the operationalization of control used in this dissertation. Control was operationalized as 'attempting to guide own cognition, behavior, attention, and motivation based on monitoring'; an example from a think aloud protocol coded as control is "*Now I will try and keep the temperature constant while increasing the mass.*". On the one hand, this operationalization could indicate that students rarely verbalized their intentions explicitly and instead they often simply performed the action (i.e., increasing the mass and keeping the temperature constant). Performing the action in this example would be coded as evidence generation instead. Nelson and Narens (1990) define metacognition as an interrelated system in which monitoring and control simultaneously affect each other and the boundaries between the two processes are not always clear. On the other hand, it is possible that monitoring (which was related to scientific reasoning activities) is tightly connected with control and the two cannot be disentangled. The difficulty of separating students' behavioral application of scientific reasoning activities from control might explain these findings. These explanations imply that the lack of association between control and scientific reasoning processes might be a statistical artefact resulting from the operationalization and coding of these constructs or the lack of verbalization of control processes. These potential explanations require further empirical investigation using better operationalizations and measurement of control processes.

Self-regulation processes were not performed conjointly with the epistemic activities problem identification and drawing conclusions. As previously discussed, the design of the inquiry tasks had already identified the problem, and thus, students did not have to engage in this epistemic activity extensively. Therefore, one explanation for this finding could be that, because the problem was already identified, students did not have to monitor and regulate this epistemic activity. Future investigations, in which the problem is not identified prior to the inquiry task, are necessary to investigate the role of self-regulation processes for problem identification. The lack of conjoint use of self-regulation processes and drawing conclusions could indicate that, because students had already generated and evaluated the evidence, monitoring during this epistemic activity was not necessary anymore.

In conclusion, these findings indicate that planning and monitoring during specific scientific reasoning activities are positively associated with learning outcomes, which corroborates prior research showing that self-regulation is associated with knowledge gains from inquiry learning (Chin & Brown, 2000; Kuhn et al., 2000). Convergence of the results from the two studies provides empirical support for the theoretical claims made about the role of self-regulation for scientific reasoning processes (Andersen & Garcia-Mila, 2017; Azevedo, 2009; Pedaste et al., 2012; White & Frederiksen, 1998; White et al., 2009). A major contribution of this dissertation is that it provides an actual in-depth process-oriented investigation of the interplay between scientific reasoning and self-regulation in relation to students' hypothesis and argumentation quality. The findings illustrated in the SRLSR model indicate that self-regulation processes need to be integrated into current theoretical frameworks of scientific reasoning (Fischer et al., 2014; Klahr & Dunbar, 1988) and inquiry learning (de

Jong & Lazonder, 2014; Pedaste et al., 2015). Quintana et al. (2005) proposed a framework that incorporates metacognition during online information search and synthesis. Findings from this dissertation corroborate prior research regarding the role of metacognitive processes and extends it to the context of computer-supported inquiry learning in science education.

3. Educational Implications

The increased attention in science education on improving scientific reasoning and argumentation has resulted in the increased implementation of active learner-centered pedagogies. The increased use of (computer-supported) inquiry learning requires further investigation of factors that contribute to its effectiveness for teaching scientific reasoning and argumentation. This dissertation investigated three factors related to students' engagement, their learning processes, and learning outcomes – 1) cognitive and motivational prerequisites, 2) self-regulated learning processes, 3) effectiveness of guidance supporting the integration of self-regulation and scientific reasoning. The findings of these studies provide important implications for future educational research and educational practice.

First, findings from Study 1 indicated that students differ regarding their cognitive and motivational prerequisites. In particular, students are not always homogeneous regarding their prior knowledge, interest, and self-concept (Kosel et al., 2021; Seidel, 2006). Both cognitive and motivational prerequisites were related to conceptual understanding and experimentation behavior. Students can be motivated while lacking prior knowledge or students might possess high levels of prior knowledge while being unmotivated to engage with the task (Huber & Seidel, 2018). These types of students are likely to benefit from different types of guidance. Furthermore, the findings demonstrated that although structured tasks support students' use of the control-of-variables strategy, they do not always ensure effective learning. Therefore, finding a balance between providing students with freedom and support is fundamental and could depend on students' prior motivational and cognitive prerequisites, corroborating prior research showing that students' self-concept or prior knowledge can be boundary conditions for the effectiveness of different types of instruction in science (Richter et al., 2022; Roelle & Renkl, 2019; van Riesen et al., 2019). Findings from this dissertation extend this prior research by demonstrating the value of considering the combined effects of students' cognitive in motivational characteristics for inquiry learning. In future research, it is important to investigate the effectiveness of different guidance measures tailored to students' prior motivational and cognitive prerequisites. Another direction for future research is investigating the relationship between students' cognitive and motivational prerequisites in relation to scientific reasoning and self-regulation processes during inquiry learning.

Second, findings from this dissertation highlight the importance of considering self-regulation processes during inquiry learning. Moreover, self-regulation processes do not occur in isolation and their temporal co-occurrence with scientific reasoning processes needs to be considered. A drawback of Study 1 was that only the control-of-variables strategy was considered as a measure of experimentation skills. Findings from Study 2 demonstrated that students engage in several scientific reasoning processes (e.g., hypothesis generation, evidence evaluation, drawing conclusions) and the conjoint use of these processes is associated with higher argumentation quality. Therefore, in future research on inquiry learning, it is important to consider a wider range of scientific processes when investigating students' inquiry learning activities. Moreover, combining data on students' self-regulation and scientific reasoning processes is necessary. Previous research has continuously demonstrated the importance of providing guidance for effective inquiry learning (for a review and meta-analysis, see Lazonder & Harmsen, 2016; Zacharia et al., 2015). Prior studies on guided inquiry learning have investigated different types of guidance (e.g., process constraints, heuristics, scaffolds) that support the inquiry process and thereby scientific reasoning processes (Zacharia et al., 2015). At the same time, the need to additionally support self-regulated learning during inquiry has been highlighted (Shapiro, 2008; Zacharia et al., 2015) and some studies have focused on the effectiveness of self-regulation scaffolds (e.g., Manlove et al., 2009b). Nevertheless, the effectiveness of guidance, which integrates self-regulation and scientific reasoning, was not previously investigated. Findings from this dissertation provide evidence for the importance of considering self-regulation processes during inquiry learning (Zacharia et al., 2015) and also extend prior research by showing the effectiveness of video modeling examples for teaching the integration of self-regulation and scientific reasoning. This dissertation focused on the role of planning, monitoring, and control for scientific reasoning. In future research, the interplay between other self-regulation processes (e.g., reflection and evaluation) and scientific reasoning, needs to be investigated.

Third, findings from Study 3 demonstrated the effectiveness of video modeling examples for fostering the self-regulation of scientific reasoning processes and for improving hypothesis and argumentation quality, thereby corroborating previous research showing that video modeling is effective (Kant et al., 2017; Mulder et al., 2014). Previous research has demonstrated the effectiveness of video modeling examples for teaching scientific reasoning (e.g., Kant et al., 2017) or self-regulated learning (e.g., Raaijmakers et al., 2018a, b). Findings from Study 3 extend previous research by showing the effectiveness of video modeling examples for providing an integrated instruction of scientific reasoning and self-regulation. Furthermore, an analysis of students' learning processes extends prior research by providing evidence for the effects of video modeling examples on the process level. Video modeling examples are relatively easy to create and can be implemented in science classrooms in combination with inquiry tasks. In Study 3, no benefits of additionally providing students with metacognitive prompts were observed. It is possible that video modeling examples were sufficient for fostering self-regulation and the metacognitive prompts were unnecessary. Another explanation could be that the design of the prompts, which only asked students to rate their confidence without providing an additional written response, resulted in their ineffectiveness. In future research, the effectiveness of prompts that require generating a written response should be investigated. Another direction for improving the effectiveness of the prompts is their timing. In Study 3, the prompts appeared immediately after participants watched the video modeling examples. Providing the prompts in a delayed task might improve their effectiveness.

Last, the video modeling examples instructed students on generating a testable hypothesis. This resulted in improved hypothesis quality in the training and transfer tasks. However, the structure of argumentation was not instructed or supported and the improvement in argumentation quality did not transfer to a novel task. Providing well-structured arguments is difficult for students (McNeill, 2011; Sadler, 2004) and they can benefit from additional instruction (Osborne et al., 2004). In future research, the effectiveness of additional instruction on argumentation (for example, using video modeling examples) needs to be investigated.

4. Strengths and Limitations

The main strengths of this research pertain to the use of advanced statistical methods to combine learning processes with learning outcomes and the variety of instructional materials and samples.

First, each study reported in this dissertation used a novel statistical method that goes beyond the traditional variable-oriented approach and instead considers the conjoint influence of different factors. In Study 1, a person-oriented approach was used to investigate the combined role of cognitive and motivational variables which accounts for students' individual differences. Students that are homogeneous regarding motivation and cognition (e.g., knowledgeable and motivated) can be identified, as well as students with heterogeneous scores on these variables (e.g., knowledgeable and unmotivated). Employing a person-oriented approach can provide insights into the effectiveness of instruction for diverse student samples which in turn can inform instructional design that is adaptive to students' individual prerequisites (Hammer et al., 2021; Kosel et al., 2021; Seidel, 2006).

Studies 2 and 3 focused extensively on students' learning processes in relation to their learning outcomes, in contrast with more traditional approaches that solely focus on the effectiveness of instruction on students' learning outcomes. Moreover, this dissertation focused on the co-occurrences and sequential patterns of learning processes, in contrast with prior research often considering only the isolated frequencies of learning processes (e.g., Kant et al., 2017; Liu et al., 2022; Mulder et al., 2014). Two methods were used to investigate the co-occurrences and sequential patterns of self-regulation and scientific reasoning processes – epistemic network analysis (ENA) and process mining. In particular, Study 2 provided insights into the interplay of self-regulation and scientific reasoning processes using ENA – a novel statistical method that considers the temporal co-occurrences of learning processes. Findings from Study 2 were used to design instruction that supports the interplay of self-regulation and scientific reasoning. The effectiveness of the instruction on students' learning processes and outcomes was tested in Study 3.

Findings from Study 2 were based on an exploratory analysis that could not provide causal claims about the interplay of scientific reasoning and self-regulation processes and argumentation quality. However, modeling the interplay in the instruction in Study 3 and investigating the intervention's effects on the learning process and outcomes level provided a basis for causal claims about the benefits of supporting the interplay of self-regulation and scientific reasoning for argumentation and hypothesis quality. Moreover, learning processes in Study 3 were modeled using two methods – ENA and process mining – that can complement the drawbacks associated with each method. While ENA can statistically compare the strength of individual relationships, information about the specific sequence is not available. This information is obtained from the process mining models. Therefore, an integration of the findings from both methods provides an in-depth analysis of the role of self-regulation during scientific reasoning and its impact on learning outcomes. Furthermore, an integration of the converging findings from Study 2 and 3 resulted in the SRLSR theoretical model (see Figure 19) illustrating the relationship between specific self-regulation and scientific reasoning processes. The findings illustrated in the SRLSR model provide the basis for future research investigating the role of other self-regulation processes (e.g., reflection) and other scientific reasoning processes (e.g., questioning, problem identification) and for designing new guidance tools that can support their interplay during inquiry learning.

The second strength of this dissertation relates to the use of a variety of instructional materials in the different studies. The inquiry simulations came from two domains in science education – biology and physics. In biology, I investigated students' scientific reasoning and argumentation regarding the topics of photosynthesis (Studies 1 and 3) and population genetics (Study 2). In physics, students solved an inquiry task on the topic of energy conversion in a system (Study 3). Using a variety of topics and domains allows for the generalizability of the findings in this dissertation, which are not constrained to only one simulation, topic, or domain. Moreover, studies were conducted with university students (Studies 2 and 3) and secondary school students (Study 1), providing insights into the inquiry learning processes of students in different educational stages.

At the same time, there are certain limitations of the present research. The first limitation concerns the difficulties with measuring scientific reasoning ability using tests. In Study 2, prior scientific reasoning ability could not be controlled for because of the low reliability of the scales that were used. Study 3 presented a different issue with scientific reasoning ability tests. Identical items were used in the pre- and posttest to investigate whether students improved in scientific reasoning ability. However, data from screen recordings and time on task indicated that students simply provided identical answers in the posttest without attempting to solve the questions again. Because of the rather short duration of the experiment (1 hour and 30 minutes), students appear to have remembered their answers. The alternative of using different items in the pre- and posttest presents difficulties with accurately measuring learning gains. Consequently, scores on the scientific reasoning ability test did not provide meaningful information about the impact of the instruction on students' scientific reasoning ability. Furthermore, scientific reasoning is not a single skill, rather it has multiple components (i.e., epistemic activities, Fischer et al., 2014). Therefore, traditional measures of internal consistency (e.g., Cronbach's alpha) might not be suitable for assessing complex skills like scientific reasoning (cf. Stadler et al., 2021). This dissertation has also demonstrated the value of using process measures to identify students' scientific reasoning process, which I argue are more meaningful measures of scientific reasoning. Nevertheless, these studies could not establish a relationship between scientific reasoning ability and scientific reasoning processes. In future research, tests of scientific reasoning, which can relate specific scientific reasoning competencies (e.g., evidence generation) to process measures of scientific reasoning, should be designed. Then, the predictive ability of such tests for actual inquiry behavior can be assessed and the relationship between scientific reasoning ability and processes can be investigated more reliably.

The second limitation of this dissertation pertains to the trade-off between collecting and analyzing fine-grained process data that provides more insight into students' learning processes and instructional effectiveness. On the one hand, the use of screen recordings and think aloud protocols provides meaningful information about students' cognition, metacognition, and behavior. On the other hand, the collection and analysis of such data require a large time investment. Concurrent think aloud protocols require data collection in individual sessions and prior studies are often based on rather smaller sample sizes (e.g., Bannert & Mengelkamp, 2008; Bannert et al., 2014; Study 2). However, the findings in this dissertation are based on an in-depth analysis of 157 participants (Studies 2 and 3) using concurrent think aloud protocols. In addition to think aloud protocols, screen recordings were used in all three studies ($N = 267$ screen videos) to obtain information about students' experimentation behavior and scientific reasoning processes. A combination of both methods provided a first attempt to conjointly investigate self-regulation and scientific reasoning processes. The findings from these studies provide important insight into students' learning processes, rather than solely considering learning outcomes. Nevertheless, the increased demands regarding time and effort invested into conducting this type of research need to be considered.

Last, the present studies focused more on investigating the relationship between scientific reasoning and self-regulation processes in relation to argumentation quality (Studies 2 and 3) and less on conceptual understanding (Study 1) as a measure of learning outcomes. Although Study 1 provided evidence regarding the combined role of students' individual prerequisites for conceptual understanding, no evidence about how students' self-regulation and scientific reasoning processes relate to conceptual understanding was provided. Furthermore, it could not be established whether video modeling examples are also effective in supporting conceptual understanding, in addition to hypothesis and argumentation quality. However, previous research has shown that video modeling was effective for supporting conceptual understanding (Kant et al., 2017).

5. Conclusion

This dissertation provides an in-depth temporal investigation of students' learning processes during inquiry learning – an instructional method that is increasingly implemented in science education to enhance scientific reasoning and argumentation. An integration of the findings from the three studies in the SRLSR model revealed the association between students' individual prerequisites, self-regulation processes, and scientific reasoning processes on learning outcomes.

The studies in this dissertation demonstrate that both students' prior motivational and cognitive prerequisites need to be considered during inquiry learning. Furthermore, to better understand inquiry learning, the interplay between students' self-regulation and scientific reasoning processes should also be considered. Evidence for the importance of this interplay was provided by distinguishing students' argumentation quality based on their conjoint use of self-regulation and scientific reasoning processes. Further causal evidence for this interplay was provided in an experimental study showing the effectiveness of an integrated instruction of self-regulation and scientific reasoning for improving not only students' learning outcomes but also their conjoint use of these processes. The findings from this dissertation rely on the use of advanced statistical methods necessary for providing an in-depth temporal insight into the interplay of scientific reasoning and self-regulation processes. These findings highlight the necessity of considering students' cognitive and metacognitive processes and extending research on inquiry learning beyond focusing on learning outcomes.

X. References

Abd-El-Khalick, F., BouJaoude, S., Duschl, R., Lederman, N. G., Mamlok-Naaman, R., Hofstein, A., Niaz, M., Treagust, D., & Tuan, H. (2004). Inquiry in science education: International perspectives. *Science Education*, *88*(3), 397–419. https://doi.org/10.1002/sce.10118

Alexander, P. A., Murphy, P. K., Woods, B. S., Duhon, K. E., & Parker, D. (1997). College instruction and concomitant changes in students' knowledge, interest, and strategy use: A study of domain learning. *Contemporary Educational Psychology, 22*, 125–146. https://doi.org/10.1006/ceps.1997.0927

Alfieri, L., Brooks, P. J., Aldrich, N. J., & Tenenbaum, H. R. (2011). Does discovery-based instruction enhance learning? *Journal of Educational Psychology*, *103*(1), 1–18. https://doi.org/10.1037/a0021017

Amsel, E., Klaczynski, P. A., Johnston, A., Bench, S., Close, J., Sadler, E., & Walker, R. (2008). A dual-process account of the development of scientific reasoning: The nature and development of metacognitive intercession skills. *Cognitive Development*, *23*(4), 452–471. https://doi.org/10.1016/j.cogdev.2008.09.002

Andersen, C., & Garcia-Mila, M. (2017). Scientific reasoning during inquiry. In K. S. Taber & B. Alpan (Eds.), *Science education. new directions in mathematics and science education* (pp. 105–117). Rotterdam: Sense. https://doi.org/10.1007/978-94-6300-749-8_8

Arastoopour, G., Shaffer, D. W., Swiecki, Z., Ruis, A. R., & Chesler, N. C. (2016). Teaching and assessing engineering design thinking with virtual internships and epistemic network analysis. *International Journal of Engineering Education*, *32*(3B), 1492–1501. https://doi.org/10.18260/p.24016

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. (2009). Theoretical, conceptual, methodological, and instructional issues in research on metacognition and self-regulated learning: A discussion. *Metacognition and Learning*, *4*(1), 87–95. https://doi.org/10.1007/s11409-009-9035-7

Azevedo, R., Guthrie, J. T., & Seibert, D. (2004). The role of self-regulated learning in fostering students' conceptual understanding of complex systems with hypermedia. *Journal of Educational Computing Research*, *30*(1–2), 87–111. https://doi.org/10.2190/dvwx-gm1t-6thq-5wc7

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., 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.

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., Taub, M., & Mudrick, N. V. (2017). Understanding and reasoning about real-time cognitive, affective, and metacognitive processes to foster self-regulation with advanced learning technologies. In D. Schunk & J. A. Greene (Eds.), *Handbook of self-regulation of learning and performance* (pp. 254–270). New York: Routledge.

Azevedo, R., & Witherspoon, A. M. (2009). Self-regulated use of hypermedia. In A. Graesser, J. Dunlosky & D. Hacker (Eds.), *Handbook of metacognition in education* (pp. 319–339). Mahwah, NJ: Erlbaum.

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. http://doi.org/10.1002/acp.3008

Baker, R., Walonoski, J., Heffernan, N., Roll, I., Corbett, A., & Koedinger, K. (2008). Why students engage in "gaming the system" behavior in interactive learning environments. *Journal of Interactive Learning Research*, *19*(2), 185–224.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs: Prentice Hall.

REFERENCES

Bannert, M. (2009). Supporting self-regulated hypermedia learning through prompts: A discussion. *Zeitschrift Fur Padagogische Psychologie*, *23*(2), 139–145. https://doi.org/10.1024/1010-0652.23.2.139

Bannert, M., & Mengelkamp, C. (2008). Assessment of metacognitive skills by means of instruction to think aloud and reflect when prompted. Does the verbalisation method affect learning?. *Metacognition and Learning*, *3*(1), 39–58. https://doi.org/10.1007/s11409-007-9009-6

Bannert, M., & Reimann, P. (2012). Supporting self-regulated hypermedia learning through prompts. *Instructional Science*, *40*(1), 193–211. https://doi.org/10.1007/s11251-011-9167-4

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

Bauer, E., Fischer, F., Kiesewetter, J., Shaffer, D. W., Fischer, M. R., Zottmann, J. M., & Sailer, M. (2020). Diagnostic activities and diagnostic practices in medical education and teacher education: An interdisciplinary comparison. *Frontiers in Psychology, 11*. https://doi.org/10.3389/fpsyg.2020.562665

Becker, S., Klein, P., Gößling, A., & Kuhn, J. (2020). Using mobile devices to enhance inquiry-based learning processes. *Learning and Instruction*, *69*, 101350. https://doi.org/10.1016/j.learninstruc.2020.101350

Bell, P., & Linn, M. C. (2000). Scientific arguments as learning artifacts: Designing for learning from the web with KIE. *International Journal of Science Education*, *22*(8), 797–817. https://doi.org/10.1080/095006900412284

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), 183–197.

Bernacki, M. L. (2018). Trace data and their analysis. In D. H. Schunk, & J. A. Greene (Eds.). *Handbook of self-regulated learning and performance* (2nd ed., pp. 370–387). New York: Routledge.

REFERENCES

Blake, C., & Scanlon, E. (2007). Reconsidering simulations in science education at a distance: Features of effective use. *Journal of Computer Assisted Learning*, *23*(6), 491–502. https://doi.org/10.1111/j.1365-2729.2007.00239.x

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

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

Brod, G. (2021). Toward an understanding of when prior knowledge helps or hinders learning. *npj Science of Learning*, *6*(1), 1–3. https://doi.org/10.1038/s41539-021-00103-w

Bryant, P., Nunes, T., Hillier, J., Gilroy, C., & Barros, R. (2015). The importance of being able to deal with variables in learning science. *International Journal of Science and Mathematics Education*, *13*(1), 145–163. http://doi.org/10.1007/s10763-013-9469-x

Bumbacher, E., Salehi, S., Wieman, C., & Blikstein, P. (2018). Tools for science inquiry learning: Tool affordances, experimentation strategies, and conceptual understanding. *Journal of Science Education and Technology, 27*, 215–235. https://doi.org/10.1007/s10956-017-9719-8

Busey, T. A., Tunnicliff, 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<109::aid-tea3>3.0.co;2-7

Chung, G. K. W. K., de Vries, F. L., Cheak, A. M., Stevens, R. H., & Bewley, W. L. (2002). Cognitive process validation of an online problem solving assessment. *Computers in Human Behavior, 18*(6), 669–684. https://doi.org/10.1016/s0747-5632(02)00023-7

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/10.1037/0022-0663.79.4.347

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 CSCL 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. (1991). Learning and instruction with computer simulations. *Education and Computing*, *6*(3–4), 217–229. https://doi.org/10.1016/0167-9287(91)80002-f

de Jong, T. (2006). Technological advances in inquiry learning. *Science*, *312*, 532–533. https://doi.org/10.1126/science.1127750

de Jong, T. (2019). Moving towards engaged learning in STEM domains; there is no simple answer, but clearly a road ahead. *Journal of Computer Assisted Learning, 35*(2), 153–167. https://doi.org/10.1111/jcal.12337

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: Cambridge University Press.

de Jong, T., & Njoo, M. (1992). Learning and instruction with computer simulations: Learning processes involved. In E. De Corte, M. Linn, H. Mandl, & L. Verschaffel (Eds.), *Computer-based learning environments and problem solving* (pp. 411–427). Berlin: Springer https://doi.org/10.1007/978-3-642-77228-3_19

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 deLeeuw, N., & Chi, M. T. H. (2003). Self-explanation: enriching a situation model or repairing a domain model? In G. Sinatra & P. Pintrich (Eds.), *Intentional conceptual change* (pp. 55–78). Mahwah, NJ: Erlbaum.

Dent, A. L., & Koenka, A. C. (2016). The relation between self-regulated learning and academic achievement across childhood and adolescence: A meta-analysis. *Educational Psychology Review, 28*(3), 425–474. https://doi.org/10.1007/s10648-015-9320-8

diSessa, A. A. (1988). *Knowledge in pieces.* In G. Forman & P. B. Pufall (Eds.), *The Jean Piaget symposium series. Constructivism in the computer age* (p. 49–70). Lawrence Erlbaum Associates, Inc.

diSessa, A. A. (2019). A friendly introduction to "Knowledge in Pieces": Modeling types of knowledge and their roles in learning. In Kaiser G., Presmeg, N. (Eds.), *Compendium for Early Career Researchers in Mathematics Education*, (pp. 245–264). ICME-13 Monograhs. Springer, Cham. https://doi.org/10.1007/978-3-030-15636-7_11

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

Duncan, R. G. (2007). The role of domain-specific knowledge in generative reasoning about complicated multileveled phenomena. *Cognition and Instruction, 25*(4), 271–336. https://doi.org/10.1080/07370000701632355

Duschl, R. A., & Osborne, J. (2002). Supporting and promoting argumentation discourse in science education. *Studies in Science Education, 38*, 39–72. https://doi.org/10.1080/03057260208560187

Engelmann, K., & Bannert, M. (2021). Analyzing temporal data for understanding the learning process induced by metacognitive prompts. *Learning and Instruction, 72*. https://doi.org/10.1016/j.learninstruc.2019.05.002

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

Ertmer, P. A., & Newby, T. J. (1996). The expert learner: Strategic, self-regulated, and reflective. *Instructional Science*, *24*(1), 1–24. https://doi.org/10.1007/bf00156001

Everitt, B. S., Landau, S., Morven, L., & Stahl, D. (2011). *Cluster Analysis* (5th edition). West Sussex, UK: John Wiley & Sons.

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., Neuhaus, B., Dorner, B., Pankofer, S., Fischer, M., Strijbos, J. W., Heene, M., & 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

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences, 111*(23), 8410–8415. https://doi.org/10.1073/pnas.1319030111

Freidenreich, H. B., Duncan, R. G., & Shea, N. (2011). Exploring middle school students' understanding of three conceptual models in genetics. *International Journal of Science Education, 33*(17), 2323–2349. https://doi.org/10.1080/09500693.2010.536997

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

Glynn, S. M., Bryan, R. R., Brickman, P., & Armstrong, N. (2015). Intrinsic motivation, self-efficacy, and interest in science. In K. A. Renninger, M. Nieswandt, S. Hidi (Eds.), *Interest in mathematics and science learning* (pp. 189–202), Washington: AERA. https://doi.org/10.3102/978-0-935302-42-4_11

Gobert, J. D., Sao Pedro, M. A., Baker, R. S., Toto, E., & Montalvo, O. (2012). Leveraging educational data mining for real-time performance assessment of scientific inquiry skills within microworlds. *Journal of Educational Data Mining*, *4*(1), 104–143. https://doi.org/10.5281/zenodo.3554646

Goldman, S. R. (2003). Learning in complex domains: When and why do multiple representations help? *Learning and Instruction*, *13*(2), 239–244. https://doi.org/10.1016/S0959-47520200023-3

Greene, J. A. (2018). *Self-regulation in education.* New York, NY: Routledge.

Greene, J. A., Anderson, J. L., O'Malley, C. E., & Lobczowksi, N. G. (2018). Fostering self-regulated science inquiry in physical sciences. In M. K. DiBenedetto (Ed.), *Connecting self-regulated learning and performance with instruction across high school content areas* (pp. 163–183). Springer International Publishing. https://doi.org/10.1007/978-3-319-90928-8_6

Greene, J. A., & Azevedo, R. (2010). The measurement of learners' self-regulated cognitive and metacognitive processes while using computer-based learning environments. *Educational Psychologist, 45*(4), 203–209. https://doi.org/10.1080/00461520.2010.51593

Greene, J. A., Deekens, V. M., Copeland, D. Z., & Yu, S. (2018). Capturing and modeling self-regulated learning using think-aloud protocols. In D. H. Schunk, & J. A. Greene (Eds.). *Handbook of self-regulation of learning and performance* (pp. 323–337). (2nd ed.). New York, NY: Routledge.

Greene, J. A., Moos, D. C., Azevedo, R., & Winters, F. I. (2008). Exploring differences between gifted and grade-level students' use of self-regulatory learning processes with hypermedia. *Computers & Education, 50*(3), 1069–1083. https://doi.org/10.1016/j.compedu.2006.10.004

Greene, J. A., Moos, D. C., & Azevedo, R. (2011). Self-regulated learning with computer-based learning environments. *New Directions for Teaching and Learning, 126*, 107–115. https://doi.org/10.1002/tl.449

Greene, J. A., Plumley, R. D., Urban, C. J., Bernacki, M. L., Gates, K. M., Hogan, K. A., Demetriou, C., & Panter, A. T. (2021). Modeling temporal self-regulatory processing in a higher education biology course. *Learning and Instruction, 72,* 101201. https://doi.org/10.1016/j.learninstruc.2019.04.002

Großmann, N., & Wilde, M. (2019). Experimentation in biology lessons: guided discovery through incremental scaffolds. *International Journal of Science Education, 41*(6), 759–781. https://doi.org/10.1080/09500693.2019.1579392

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

Hagemans, M. G., van der Meij, H., & de Jong, T. (2013). The effects of a concept map-based support tool on simulation-based inquiry learning. *Journal of Educational Psychology, 105*(1), 1–24. https://doi.org/10.1037/a0029433

Hammer, M., Göllner, R., Scheiter, K., Fauth, B., & Stürmer, K. (2021). For whom do tablets make a difference? Examining student profiles and perceptions of instruction with tablets. *Computers & Education*, *166*, 104147. https://doi.org/10.1016/j.compedu.2021.104147

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., Upmeier 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

Haskel-Ittah, M., Duncan, R. G., Vázquez-Ben, L., & Yarden, A. (2019). Reasoning about genetic mechanisms: Affordances and constraints for learning. Journal of *Research in Science Teaching, 57*(3), 342–367. https://doi.org/10.1002/tea.21595

Hattie, J. A. C., & Donoghue, G. M. (2016). Learning strategies: A synthesis and conceptual model. *npj Science of Learning*, *1*, 16013. https://doi.org/10.1038/npjscilearn.2016.13

Hayenga, A. O., & Corpus, J. H. (2010). Profiles of intrinsic and extrinsic motivations: A person-centered approach to motivation and achievement in middle school. M*otivation and Emotion*, *34*, 371–383. http://doi.org/10.1007/s11031-010-9181-x

Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford Publications.

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. (2001). Interest, reading, and learning: Theoretical and practical considerations. *Educational Psychology Review*, *13*(3), 191–209. https://doi.org/10.1023/a:1016667621114

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

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

Hmelo, C. E., Nagarajan, A., & Roger, S. (2000). Effects of high and low prior knowledge on construction of a joint problem space. *Journal of Experimental Education, 69*, 36–56. https://doi.org/10.1080/00220970009600648

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, *42*(2), 99–107. https://doi.org/10.1080/00461520701263368

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

Howard, M. C., & Hoffman, M. E. (2018). Variable-centered, person-centered, and person-specific approaches: Where theory meets the method. *Organizational Research Methods*, *21*(4), 846–876. https://doi.org/10.1177/1094428117744021

Huang, C. (2011). Self-concept and academic achievement: A meta-analysis of longitudinal relations. *Journal of School Psychology, 49*, 505–528. http://doi.org/10.1016/j.jsp.2011.07.001

Huber, S. A., & Seidel, T. (2018). Comparing teacher and student perspectives on the interplay of cognitive and motivational-affective student characteristics. *PloS One, 13*(8), Article e0200609. https://doi.org/10.1371/journal.pone.0200609

Husnaini, S. J., & Chen, S. (2019). Effects of guided inquiry virtual and physical laboratories on conceptual understanding, inquiry performance, scientific inquiry self-efficacy, and enjoyment. *Physical Review Physics Education Research, 15*(1), 010119. https://doi.org/10.1103/physrevphyseducres.15.010119

Jacob, L., Lachner, A., & Scheiter, K. (2022). Do school students' academic self-concept and prior knowledge constrain the effectiveness of generating technology-mediated explanations?. *Computers & Education*, *182*, 104469. https://doi.org/10.1016/j.compedu.2022.104469

Jansen, M., Scherer, R., & Schroeders, U. (2015). Students' self-concept and self-efficacy in the sciences: Differential relations to antecedents and educational outcomes. *Contemporary Educational Psychology*, *41*, 13–24. https://doi.org/10.1016/j.cedpsych.2014.11.002

Jansen, M., Schroeders, U., & Lüdtke, O. (2014). Academic self-concept in science: Multidimensionality, relations to achievement measures, and gender differences. *Learning and Individual Differences, 30*, 11–21. https://doi.org/10.1016/j.lindif.2013.12.003

Jiménez-Aleixandre, M. P., Rodríguez, A. B., & Duschl, R. A. (2000). "Doing the lesson" or "doing science": Argument in high school genetics. *Science Education, 84*, 757–792. https://doi.org/10.1002/1098-237X(200011)84:6<757::AID-SCE5>3.0.CO;2-F

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review, 19*, 509–539. https://doi.org/10.1007/s10648-007-9054-3

Kanari, Z., & Millar, R. (2004). Reasoning from data: how students collect and interpret data in science investigations. *Journal of Research in Science Teaching, 41*(7), 748–769. https://doi.org/10.1002/tea.20020

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

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*(2), 75–86.

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, *12*, 1–48. http://doi.org/10.1016/0364-0213(88)90007-9

KMK. (2020). *Bildungsstandards im Fach Biologie für die Allgemeine Hochschulreife: Beschluss der Kultusministerkonferenz vom 18.06.2020*.

Kober, N. (2015). *Reaching students: What research says about effective instruction in undergraduate science and engineering. Board on science education, division of behavioral and social sciences and education.* Washington, DC: The National Academies Press.

Koenen, J. (2014). *Entwicklung und Evaluation von experimentgestützten Lösungsbeispielen zur Förderung naturwissenschaftlich – experimenteller Arbeitsweisen* [Development and evaluation of experiment-supported worked-examples to foster scientific inquiry]. Berlin: Logos.

Kosel, C., Wolter, I., & Seidel, T. (2021). Profiling secondary school students in mathematics and German language arts using learning-relevant cognitive and motivational-affective characteristics. *Learning and Instruction*, *73*, 101434. https://doi.org/10.1016/j.learninstruc.2020.101434

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). New York, NJ: 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., & Upmeier zu Belzen, A. (2020). Measuring scientific reasoning competencies – Multiple aspects of validity. In: Zlatkin-Troitschanskaia, O., Pant, H., Toepper, M. & Lautenbach, C. (Eds.), *Student Learning in German Higher Education* (pp.261–280). Springer.

Krüger, J. T., Höffler, T. N., Wahl, M., Knickmeier, K., & Parchmann, I. (2022). Two comparative studies of computer simulations and experiments as learning tools in school and out-of-school education. *Instructional Science*, *50*(2), 169–197. https://doi.org/10.1007/s11251-021-09566-1

Kuhn, D. (1991). *The skills of argument*. Cambridge University Press.

Kuhn, D. (1993). Science as argument: Implications for teaching and learning scientific thinking. *Science Education*, *77*(3), 319–337. https://doi.org/10.1002/sce.3730770306

Kuhn, D. (2010). What is scientific thinking and how does it develop? In U. Goswami (Ed.), *Blackwell handbook of childhood cognitive development* (2nd ed., pp. 497–523). Oxford: Blackwell Publishing. http://doi.org/10.1002/9780470996652.ch17

Kuhn, D. (2021). Metacognition matters in many ways. *Educational Psychologist*, 1–14. https://doi.org/10.1080/00461520.2021.1988603

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/s1532690xci1804_3

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

Land, S. M., & Greene, B. A. (2000). Project-based learning with the world wide web: A qualitative study of resource integration. *Educational Technology Research and Development*, *48*(1), 45–66. https://doi.org/10.1007/bf02313485

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159. https://doi.org/10.2307/2529310

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

Lazonder, A. W., & Kamp, E. (2012). Bit by bit or all at once? Splitting up the inquiry task to promote children's scientific reasoning. *Learning and Instruction, 22*, 458–464. http://doi.org/10.1016/j.learninstruc.2012.05.005

Lazonder, A. W., Wilhelm, P., & Hagemans, G. M. (2008). The influence of domain knowledge on strategy use during simulation-based inquiry learning. *Learning and Instruction*, *18*, 580–592. https://doi.org/10.1016/j.learninstruc.2007.12.001

Lim, L., Bannert, M., Graaf, J., Molenaar, I., Fan, Y., Kilgour, J., Moore, J., & Gašević, D. (2021). Temporal assessment of self-regulated learning by mining students' think-aloud protocols. *Frontiers in Psychology, 12*, 749749. https://doi.org/10.3389/fpsyg.2021.749749

Lin, X. (2001). Designing metacognitive activities. *Educational Technology Research and Development, 49*(2), 23–40. https://doi.org/10.1007/bf02504926

Liu, C. C., Wen, C. T., Chang, H. Y., Chang, M. H., Lai, P. H., Fan Chiang, S. H., Yang, C. W. & Hwang, F. K. (2022). Augmenting the effect of virtual labs with" teacher demonstration" and" student critique" instructional designs to scaffold the development of scientific literacy. *Instructional Science*, 1–31. https://doi.org/10.1007/s11251-021-09571-4

López-Rubio, E., Palomo, E. J., & Ortega-Zamorano, F. (2018). Unsupervised learning by cluster quality optimization. *Information Sciences, 436*, 31–55.

Magno, C. (2011). Assessing the relationship of scientific thinking, self-regulation in research, and creativity in a measurement model. *International Journal of Research & Review*, *6*(1), 17–47.

Mäeots, M., Siiman, L., Kori, K., Eelmets, 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. (2006). Regulative support for collaborative scientific inquiry learning. *Journal of Computer Assisted Learning*, *22*(2), 87–98. https://doi.org/10.1111/j.1365-2729.2006.00162.x

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. (2009a). Collaborative versus individual use of regulative software scaffolds during scientific inquiry learning. *Interactive Learning Environments, 17*(2), 105–117. https://doi.org/10.1080/10494820701706437

Manlove, S., Lazonder, A. W., & de Jong, T. (2009b). 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 1.6.0) [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

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? *American Psychologist, 59*(1), 14–19. https://doi.org/10.1037/0003-066x.59.1.14

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. http://doi.org/10.1037/0278-7393.21.5.1263

McCarthy, K. S., & McNamara, D. S. (2021). The multidimensional knowledge in text comprehension framework. *Educational Psychologist*, *56*(3), 1–19. https://doi.org/10.1080/00461520.2021.1872379

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

Molenaar, I. (2014). Advances in temporal analysis in learning and instruction. *Frontline Learning Research, 2*(4), 15–24. https://doi.org/10.14786/flr.v2i4.118

Möller, J., Zitzmann, S., Helm, F., Machts, N., & Wolff, F. (2020). A meta-analysis of relations between achievement and self-concept. *Review of Educational Research*, *90*(3), 376–419. https://doi.org/10.3102/0034654320919354

Mulder, Y. G., Lazonder, A. W., & de Jong, T. (2009). Finding out how they find it out: An empirical analysis of inquiry learners' need for support. *International Journal of Science Education, 32*(15), 2033–2053. https://doi.org/10.1080/09500690903289993

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. http://doi.org/10.1016/j.learninstruc.2013.08.001

National Academies of Sciences, Engineering, and Medicine. (2016). *Science literacy: Concepts, contexts, and consequences*. Washington, DC: The National Academies Press. https://doi.org/10.17226/23595

National Academies of Sciences, Engineering, and Medicine. (2021). *Call to action for science education: Building opportunity for the future*. Washington, DC: The National Academies Press. https://doi.org/10.17226/26152

National Research Council. (1996). *National science education standards*. Washington, DC: National Academy Press.

National Research Council. (2000). *Inquiry and the national science education standards: A guide for teaching and learning*. Washington, DC: National Academies Press.

National Research Council. (2013). *Next generation science standards: For states, by states*. Washington, DC: The National Academies Press.

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). New York, NY: Academic Press.

NGSS Lead States. (2013). *Next generation science standards: For states, by states*. The National Academies Press.

Njoo, M., & de Jong, T. (1993). Exploratory learning with a computer simulation for control theory: Learning processes and instructional support. *Journal of Research in Science Teaching*, *30*(8), 821–844. https://doi.org/10.1002/tea.3660300803

OECD. (2003). *Student engagement at school: A sense of belonging and participation.* Paris: OECD.

OECD. (2013). *Trends shaping education*. Paris: OECD. https://doi.org/10.1787/trends_edu-2013-en

OECD. (2015). *PISA 2015 released field trial cognitive tiems.* Retrieved from: http://www.oecd.org/pisa/pisaproducts/PISA2015-Released-FT-Cognitive-Items.pdf

OECD. (2016). *PISA 2015 Results: Excellence and equity in education*. https://doi.org/10.1787/9789264266490-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. (2010). Arguing to learn in science: The role of collaborative, critical discourse. *Science, 328*(5977), 463–466. https://doi.org/10.1126/science.1183944

Osborne, J. (2013). The 21st century challenge for science education: Assessing scientific reasoning. *Thinking Skills and Creativity*, *10*, 265–279. http://doi.org/10.1016/j.tsc.2013.07.006

Osborne, J. (2018). Styles of scientific reasoning. In F. Fischer, C. A. Chinn, K. Engelmann, & J. Osborne (Eds.), *Scientific reasoning and argumentation* (pp. 162–186). New York, NY: Routledge. https://doi.org/10.4324/9780203731826-9

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. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*, *8*(422). https://doi.org/10.3389/fpsyg.2017.00422

Panadero, E., & Alonso-Tapia, J. (2014). How do students self-regulate? Review of Zimmerman's cyclical model of self-regulated learning. *Anales de psicologia*, *30*(2), 450–462. http://doi.org/10.6018/analesps.30.2.167221

Pedaste, M., Mäeots, M., Leijen, Ä., & Sarapuu, T. (2012). Improving students' inquiry skills through reflection and self-regulation scaffolds. *Technology, Instruction, Cognition and Learning*, *9*(1–2), 81–95. https://doi.org/10.1109/icalt.2008.239

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

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

Puustinen, M., & Pulkkinen, L. (2001). Models of self-regulated learning: A review. *Scandinavian Journal of Educational Research*, *45*(3), 269–286. https://doi.org/10.1080/00313830120074206

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. *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

Rack, O., Zahn, C. & Mateescu, M. (2018). Coding and counting – Frequency analysis for group interaction. In Brauner, E., Boos, M., & Kolbe, M. (Eds.) *The Cambridge Handbook of Group Interaction Analysis*. Cambridge, UK: Cambridge University Press.

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

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

Reimann, P. (2009). Time is precious: Variable- and event-centered 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. (1997). Learning from worked-out examples: A study on individual differences. *Cognitive Science, 21*(1), 1–29. https://doi.org/10.1207/s15516709cog2101_1

Renkl, A. (2002). Worked-out examples: Instructional explanations support learning by self-explanations. *Learning and Instruction, 12*(5), 529–556. https://doi.org/10.1016/s0959-4752(01)00030-5

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., Austin, L., Bachrach, J. E., Chau, A., Emmerson, M. S., King, B. R., Riley, K. R., Stevens, S. J. (2014). Going beyond whoa! That's cool! Achieving science interest and learning with the ICAN Intervention. In S. Karabenick & T. Urdan (Eds.), *Motivation-based learning interventions: Advances in motivation and achievement series* (Vol. 18, pp. 107–140). https://doi.org/10.1108/S0749-742320140000018003

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). New York: Cambridge University Press.

Renninger, K. A., & Hidi, S. (2011). Revisiting the conceptualization, measurement, and generation of interest. *Educational Psychologist, 46*, 168–184. https://doi.org/10.1080/00461520.2011.587723

Renninger, K. A., Ren, Y., & Kern, H. M. (2018). Motivation, engagement, and interest: "In the end, it came down to you and how you think of the problem". In F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, & P. Reimann (Eds.), *International handbook of the learning sciences* (pp. 116–126). New York: Routledge.

Richter, J., Lachner, A., Jacob, L., Bilgenroth, F., & Scheiter, K. (2022). Self-concept but not prior knowledge moderates effects of different implementations of computer-assisted inquiry learning activities on students' learning. *Journal of Computer Assisted Learning*, 1–19. https://doi.org/10.1111/jcal.12673

Rodrigues, S. (2007). Factors that influence pupil engagement with science simulations: The role of distraction, vividness, logic, instruction and prior knowledge. *Chemistry Education Research and Practice, 8*, 1–12. https://doi.org/10.1039/b6rp90016j

Roelle, J., & Renkl, A. (2019). Does an option to review instructional explanations enhance example-based learning? It depends on learners' academic self-concept. *Journal of Educational Psychology, 112*, 131–147. https://doi.org/10.1037/edu0000365

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.

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

Rutten, N., van Joolingen, W. R., & van der Veen, J. T. (2012). The learning effects of computer simulations in science education. *Computers and Education*, *58*(1), 136–153. https://doi.org/10.1016/j.compedu.2011.07.017

Sadler, T. D. (2004). Informal reasoning regarding socioscientific issues: A critical review of research. *Journal of Research in Science Teaching 41*(5), 513–536. https://doi.org/10.1002/tea.20009

Saint, J., Fan, Y., Singh, S., Gašević, D. and Pardo, A. (2021). Using process mining to analyse self-regulated learning: A systematic analysis of four algorithms. *Proceedings of the 11th International Learning Analytics and Knowledge Conference, ACM* (pp. 333–343). https://doi.org/10.1145/3448139.3448171

Sandoval, W. A. & Çam, A. (2011). Elementary children's judgments of the epistemic status of sources of justification. *Science Education, 95*(3), 383–408. https://doi.org/10.1002/sce.20426

Sandoval, W. A., & Millwood, K. (2005). The quality of students' use of evidence in written scientific explanations. *Cognition and Instruction, 23*(1), 23–55. https://doi.org/10.1207/s1532690xci2301_2

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.

Schauble, L., Glaser, R., Raghavan, K., & Reiner, M. (1991). Causal models and experimentation strategies in scientific reasoning. *Journal of the Learning Sciences, 1*, 201–238. https://doi.org/10.1207/s15327809jls0102_3

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

Scheiter, K., Gerjets, P., Vollmann, B., & Catrambone, R. (2009). The impact of learner characteristics on information utilization strategies, cognitive load experienced, and performance in hypermedia learning. *Learning and Instruction, 19*, 387–401. https://doi.org/10.1016/j.learninstruc.2009.02.004

Schiefele, U., Krapp, A., & Winteler, A. (1992). Interest as a predictor of academic achievement: A meta-analysis of research. In K. A. Renninger, S. Hidi, & A. Krapp (Eds.), *The role of interest in learning and development* (pp. 183–212). Lawrence Erlbaum Associates, Inc.

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

Schnitzler, K., Holzberger, D., & Seidel, T. (2021). All better than being disengaged: Student engagement patterns and their relations to academic self-concept and achievement. *European Journal of Psychology of Education*, *36*(3), 627–652. https://doi.org/10.1007/s10212-020-00500-6

Schraw, G. (1998). Promoting general metacognitive awareness. *Instructional Science*, *26*(1/2), 113–125. https://doi.org/10.1023/a:1003044231033

Schraw, G. (2010). Measuring self-regulation in computer-based learning environments. *Educational Psychologist*, *45*(4), 258–266. https://doi.org/10.1080/00461520.2010.515936

Schraw, G., Crippen, K. J., & Hartley, K. (2006). Promoting self-regulation in science education: Metacognition as part of a broader perspective on learning. *Research Science Education*, *36*(1–2), 111–139. https://doi.org/10.1007/s11165-005-3917-8

Schraw, G., & Moshman, D. (1995). Metacognitive theories. *Educational Psychology Review*, *7*(4), 351–371. https://doi.org/10.1007/BF02212307

Schunk, D., H. (2005). Self-regulated learning: The educational legacy of Paul R. Pintrich. *Educational Psychologist*, *40*(2), 85–94. https://doi.org/10.1207/s15326985ep4002_3

Schunk, D. H., & Greene, J. A. (Eds.). (2017). *Handbook of self-regulation of learning and performance* (2nd ed.). New York: Routledge.

Schunk, D. H., & Zimmerman, B. J. (Eds.). (1998). *Self-regulated learning: From teaching to self-reflective practice*. Guilford Press.

Schunk, D. H., & Zimmerman, B. J. (2006). Competence and control beliefs: Distinguishing the means and ends. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology* (2nd ed., pp. 349–367). Washington, DC: American Psychology Association.

Schwarz, B. B. (2009). Argumentation and learning. In N. M. Mirza & A. N. Perret-Clermont (Eds.), *Argumentation and education: Theoretical foundations and practices* (pp. 91–126). Boston, MA: Springer US. https://doi.org/10.1007/978-0-387-98125-3_4

Schwichow, M., Croker, S., Zimmerman, C., Höffler, T., & Härtig, H. (2016). Teaching the control-of-variables strategy: A meta-analysis. *Developmental Review, 39*(1), 37–63. https://doi.org/10.1016/j.dr.2015.12.001

Schwichow, M., Osterhaus, C., & Edelsbrunner, P. A. (2020). The relation between the control-of-variables strategy and content knowledge in physics in secondary school. *Contemporary Educational Psychology*, *63*, 101923. https://doi.org/10.1016/j.cedpsych.2020.101923

Seidel, T. (2006). The role of student characteristics in studying micro teaching-learning environments. *Learning Environments Research, 9*, 253–271. https://doi.org/10.1007/s10984-006-9012-x

Seufert, T. (2018). The interplay between self-regulation in learning and cognitive load. *Educational Research Review, 24*, 116–129. https://doi.org/10.1007/s11251-020-09502-9

Shaffer, D. W. (2007). *How computer games help children learn*. New York, NY: Palgrave Macmillan.

Shaffer, D. W. (2012). Models of situated action: Computer games and the problem of transfer. In C. Steinkuehler, K. Squire, & S. Barab (Eds.), Games learning, and society: *Learning and meaning in the digital age* (pp. 403–433). Cambridge, UK: Cambridge University Press.

Shaffer, D. W. (2017). *Quantitative ethnography.* Madison, WI: Cathcart Press.

Shaffer, D. W. (2018). Epistemic network analysis: Understanding learning by using big data for thick description. In F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, & P. Reimann (Eds.), *International handbook of the learning sciences* (pp. 520–531). Routledge.

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

Shaffer, D. W., & Ruis, A. R. (2017). Epistemic network analysis: A worked example of theory-based learning analytics. In C. Lang, G. Siemens, A. F. Wise, & D. Gasevic (Eds.), *Handbook of learning analytics* (pp. 175–187). Society for Learning Analytics Research.

Shapiro, A. M. (2008). Hypermedia design as learner scaffolding. *Educational Technology Research and Development, 56*, 29–44. https://doi.org/10.1007/s11423-007-9063-4

Siebert-Evenstone, A. L., Arastoopour , G., Collier, W., Swiecki, Z., Ruis, A. R., & Shaffer, D. W (2017). In search of conversational grain size: Modelling semantic structure using moving stanza windows. *Journal of Learning Analytics*, *4*(3), 123–139. https://doi.org/10.18608/jla.2017.43.7

Sinatra, G. M., & Hofer, B. (2021). *Science denial: Why it happens and what to do about it*. Oxford University Press.

Sinatra, G. M., & Taasoobshirazi, G. (2018). The self-regulation of learning and conceptual change in science: Research, theory, and educational applications. In D. H. Schunk &

J. A. Greene (Eds.), *Handbook of self-regulation of learning and performance* (2nd ed., pp. 153–165). New York, NY: Routledge.

Simonsmeier, B. A., Flaig, M., Deiglmayr, A., Schalk, L., & Schneider, M. (2021). Domain-specific prior knowledge and learning: A meta-analysis. *Educational Psychologist*, 1–24. https://doi.org/10.1080/00461520.2021.1939700

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

Smith, C. & Wenk, L. (2006). Relations among three aspects of first-year college students' epistemologies of science. *Journal of Research in Science Teaching*, *43*(8), 747–785. https://doi.org/10.1002/tea.20113

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

Stender, A., Schwichow, M., Zimmerman, C., & Härtig, H. (2018). Making inquiry-based science learning visible: The influence of CVS and cognitive skills on content knowledge learning in guided inquiry. *International Journal of Science Education*, 1–20. https://doi.org/10.1080/09500693.2018.1504346

Sullivan, S., Warner-Hillard, C., Eagan, B., Thompson, R. J., Ruis, A. R., Haines, K., Pugh, C. M, Shaffer, D. W., & Jung, H. S. (2018). Using epistemic network analysis to identify targets for educational interventions in trauma team communication. *Surgery, 163*(4), 938–943. https://doi.org/10.1016/j.surg.2017.11.009

Suthers, D. D. (2006). Technology affordances for intersubjective meaning making: A research agenda for CSCL. *International Journal of Computer-Supported Collaborative Learning, 1*(3), 315–337. https://doi.org/10.1007/s11412-006-9660-y

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York: Springer.

Tarchi, C. (2021). Effects of think-aloud on students’ multiple-documents comprehension. *Applied Cognitive Psychology, 35,* 526–537. https://doi.org/10.1002/acp.3782

Thörne, K., & Gericke, N. (2014). Teaching genetics in secondary classrooms: A linguistic analysis of teachers' talk about proteins. *Research in Science Education, 44*(1), 81–108. https://doi.org/10.1007/s11165-013-9375-9

Tobias, S. (1994). Interest, prior knowledge, and learning. *Review of Educational Research, 64*(1), 37–54. https://doi.org/10.3102/00346543064001037

Toulmin, S. (1958). *The uses of argument*. Cambridge: Cambridge University Press.

Trautwein, U., Lüdtke, O., Schnyder, I., & Niggli, A. (2006). Predicting homework effort: support for a domain-specific, multilevel homework model. *Journal of Educational Psychology, 98*, 438–456. https://doi.org/10.1037/0022-0663.98.2.438

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., 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 Gog, T., & Rummel, N. (2010). Example-based learning: Integrating cognitive and social-cognitive research perspectives. *Educational Psychology Review*, *22*, 155–174. https://doi.org10.1007/s10648-010-9134-7

van Joolingen, W. R., de Jong, T., & Dimitrakopoulou, A. (2007). Issues in computer supported inquiry learning in science. *Journal of Computer Assisted Learning*, *23*(2), 111–119. https://doi.org/10.1111/j.1365-2729.2006.00216.x

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

van Mil, M. H. W., Postma, P. A., Boerwinkel, D. J., Klaassen, K., & Waarlo, A. J. (2016). Molecular mechanistic reasoning: Toward bridging the gap between the molecular and cellular levels in life science education. *Science Education, 100*(3), 517–585. https://doi.org/10.1002/sce.21215

van Riesen, S. A. N., Gijlers, H., Anjewierden, A., & de Jong, T. (2018). The influence of prior knowledge on experiment design guidance in a science inquiry context. *International Journal of Science Education, 40*(11), 1327–1344. https://doi.org/10.1080/09500693.2018.1477263

van Riesen, S. A. N., Gijlers, H., Anjewierden, A. A., & de Jong, T. (2019). The influence of prior knowledge on the effectiveness of guided experiment design. *Interactive Learning Environments*, 1–17. https://doi.org/10.1080/10494820.2019.1631193

Veenman, M. V. J., Elshout, J. J., & Busato, V. V. (1994). Metacognitive mediation in learning with computer-based simulations. *Computers in Human Behavior*, *10*, 93–106. https://doi.org/10.1016/0747-5632(94)90031-0

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.

von Aufschnaiter, C., Erduran, S., Osborne, J., & Simon, S. (2008). Arguing to learn and learning to argue: Case studies of how students' argumentation relates to their scientific knowledge. *Journal of Research in Science Teaching, 45*, 101–131. https://doi.org/10.1002/tea.20213

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/s1532690xci1601_2

White, B. Y., & Frederiksen, J. R. (2005). A theoretical framework and approach for fostering metacognitive development. *Educational Psychologist, 40*, 211–223. https://doi.org/10.1207/s15326985ep4004_3

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). New York, NY: Routledge.

White, B. Y., Shimoda, T. A., & Frederiksen, J. R. (1999). Enabling students to construct theories of collaborative inquiry and reflective learning: computer support for metacognitive development. *International Journal of Artificial Intelligence in Education, 10*, 151–182.

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 Kurzskala 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.

Winne, P. H. (1996). A metacognitive view of individual differences in self-regulated learning. *Learning and Individual Differences, 8*(4), 327–353. http://doi.org/10.1016/S1041-6080(96)90022-9

Winne, P. H. (2001). Self-regulated learning viewed from models of information processing. In B. J. Zimmerman & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives* (pp. 153–189). Lawrence Erlbaum Associates.

Winne, P. H. (2010). Improving measurements of self-regulated learning. *Educational Psychologist, 45*(4), 267–276. https://doi.org/10.1080/00461520.2010.51715

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Hillsdale: Erlbaum.

Winne, P. H., & Perry, N. E. (2000). Measuring self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 531–566). Orlando, FL: Academic Press.

Wolters, C. A., & Taylor, D. J. (2012). A self-regulated learning perspective on student engagement. In *Handbook of research on student engagement* (pp. 635–651). Springer, Boston, MA.

Wu, L., Liu, Q., Mao, G., & Zhang, S. (2020). Using epistemic network analysis and self-reported reflections to explore students' metacognition differences in collaborative learning. *Learning and Individual Differences, 82*, 101913. https://doi.org/10.1016/j.lindif.2020.101913

Zacharia, Z. C., Manoli, C., Xenofontos, N., de Jong, T., Pedaste, M., van Riesen, S. A. N., Kamp, E. T., Mäeots, M., Siiman, L., & 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

Zhang, J., Chen, Q., Sun, Y., & Reid, D. J. (2004). Triple scheme of learning support design for scientific discovery learning based on computer simulation: Experimental research. *Journal of Computer Assisted Learning, 20*, 269–282. https://doi.org/10.1111/j.1365-2729.2004.00062.x

Zhang, L., Kirschner, P. A., Cobern, W. W., & Sweller, J. (2021). There is an evidence crisis in science educational policy. *Educational Psychology Review*. https://doi.org/10.1007/s10648-021-09646-1

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.

Zhou, M., & Winne, P. H. (2012). Modeling academic achievement by self-reported versus traced goal orientation. *Learning and Instruction, 22*(6), 413–419. https://doi.org/10.1016/j.learninstruc.2012.03.004

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). San Diego: Academic Press.

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). New York: Routledge.

Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review*, *27*(2), 172–223. http://doi.org/10.1016/j.dr.2006.12.001

Zimmerman, C., & Croker, S. (2013). Learning science through inquiry. In G. J. Feist, & M. E. Gorman (Eds.), *Handbook of the psychology of science* (pp. 49–70). New York: Springer.

Zohar, A., & Nemet, F. (2002). Fostering students' knowledge and argumentation skills through dilemmas in human genetics. *Journal of Research in Science Teaching*, *39*(1), 35–62. https://doi.org/10.1002/tea.10008

Appendix A

Online Supplementary Material Study 2

Table A1

Mean Proportions (and SDs) of Scientific Reasoning, Self-Regulation, and Generic Cognitive Strategies

	Argumentation Quality		
Process	Low (n = 15)	High (n = 15)	Total (N = 30)
Scientific Reasoning			
Problem Identification	0	.006 (.022)	.001
Hypothesis Generation			
correct	.003 (.008)	.023 (.044)	.011
incorrect	.052 (.131)	.014 (.031)	.013
Evidence Generation			
correct	.235 (.153)	.351 (.129)	.337
incorrect	.458 (.254)	.223 (.173)	.330
Evidence Evaluation			
correct	.058 (.094)	.164 (.143)	.147
incorrect	.015 (.025)	.049 (.103)	.032
Drawing Conclusions			
correct	.003 (.009)	.046 (.081)	.023
incorrect	.005 (.014)	.008 (.026)	.005
Self-Regulated Learning			
Planning	.001 (.005)	.010 (.015)	.010
Monitoring			
negative	.089 (.256)	.022 (.041)	.019
positive	0	.010 (.026)	.003
Control	.013 (.038)	.028 (.047)	.019
Generic Cognitive Strategies			
Activation of Prior Knowledge	.008 (.017)	.017 (.031)	.011
Self-Explaining			
correct	.057 (.127)	.032 (.039)	.034
incorrect	.003 (.010)	0	.004
Total number of codes	25.00 (26.72)	27.47 (18.25)	787

Note. The Total column presents the proportion of each code in the whole sample.

Table A2

Spearman Correlations Between Scientific Reasoning and Self-Regulation Processes and Argumentation Quality

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. Argumentation	1.00																
2. PI	.19	1.00															
3. HG correct	.16	.39*	1.00														
4. HG incorrect	-.14	-.10	-.004	1.00													
5. EG correct	.46*	-.16	.05	-.38*	1.00												
6. EG incorrect	-.51**	-.06	-.17	.06	-.21	1.00											
7. EE correct	.43*	.06	.11	-.19	.19	-.45*	1.00										
8. EE incorrect	.32	-.14	-.01	.13	.13	-.37*	.46*	1.00									
9. DC correct	.39*	.33	.41*	-.35	.17	-.26	.30	-.004	1.00								
10. DC incorrect	-.08	-.07	-.05	.05	-.22	-.09	.36*	.34	-.25	1.00							
11. Planning	.30	-.09	.26	-.11	.10	-.02	.22	.09	.31	-.06	1.00						
12. Monitoring (positive)	.32	.72***	.22	-.15	.08	-.24	-.09	.15	.16	-.10	-.13	1.00					
13. Monitoring (negative)	.002	-.12	-.35	-.04	-.39*	-.18	.24	.14	-.18	.43*	-.05	-.17	1.00				
14. Control	.19	-.11	.30	-.01	.05	-.13	.07	.13	.09	.19	.17	-.16	-.02	1.00			
15. APK	.10	.41*	.24	-.13	.04	-.004	.11	-.16	.29	-.21	.06	.23	-.35	.17	1.00		
16. SE (correct)	.20	.23	.25	.22	.08	-.16	.11	.03	.05	-.14	.14	.06	-.20	.16	-.05	1.00	
17. SE (incorrect)	-.08	-.05	.13	-.15	.10	.09	.02	.25	.28	-.10	.15	-.07	-.17	.08	.11	.17	1.00

Note. PI = Problem Identification, HG = Hypothesis Generation, EG = Evidence Generation, EE = Evidence Evaluation, DC = Drawing Conclusions, APK = Activation of Prior Knowledge, SE = Self-explaining. Argumentation quality was measured on a scale from 0-6.

*$p < 0.05$, **$p < .01$, ***$p < .001$

Appendix B

Online Supplementary Material Study 3

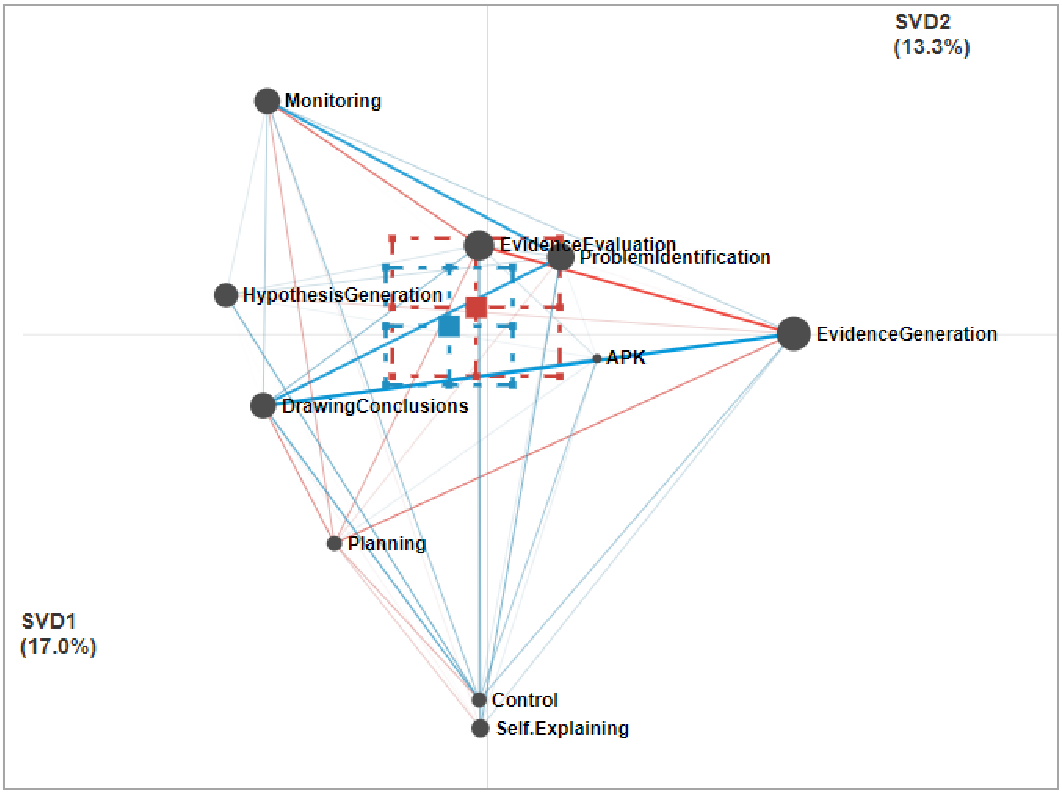
Training Task

Epistemic Network Analysis

Along the X axis, a two sample *t*-test showed that the centroid of the VM group (M = -0.16, SD = 0.78, N = 37) was not significantly different from the VMP group (M = -0.05, SD = 0.91, N = 29; $t(55.53) = 0.52$, $p = .60$, $d = 0.13$). Along the Y axis, a two sample *t*-test showed that the position of the VM group (M = 0.03, SD = 0.73, N = 37) was not significantly different from the position of the VMP group (M = 0.11, SD = 0.75, N = 29; $t(59.54) = -0.42$, $p = .67$, $d = 0.11$). These findings indicate that, in the training task, the epistemic networks of participants in the two video modeling conditions did not differ significantly from each other. From Figure B1 we can see that the networks of the two conditions did not differ substantially. We observe slightly stronger connections between *planning*, *monitoring*, *evidence generation*, and *evidence evaluation*, indicated by red lines in Figure B1. In the VM condition, we observed stronger connections between *monitoring* and *problem identification*, *evidence generation* and *drawing conclusions*, and *problem identification* and *drawing conclusions* than in the VMP condition. However, since we found no significant differences between the centroids, these differences are purely descriptive. Similar to our findings on the product level, we found no significant differences on the process level between the two video modeling conditions.

Figure B1

The Network Difference between the VMP (red) Condition and the VM (blue) Condition



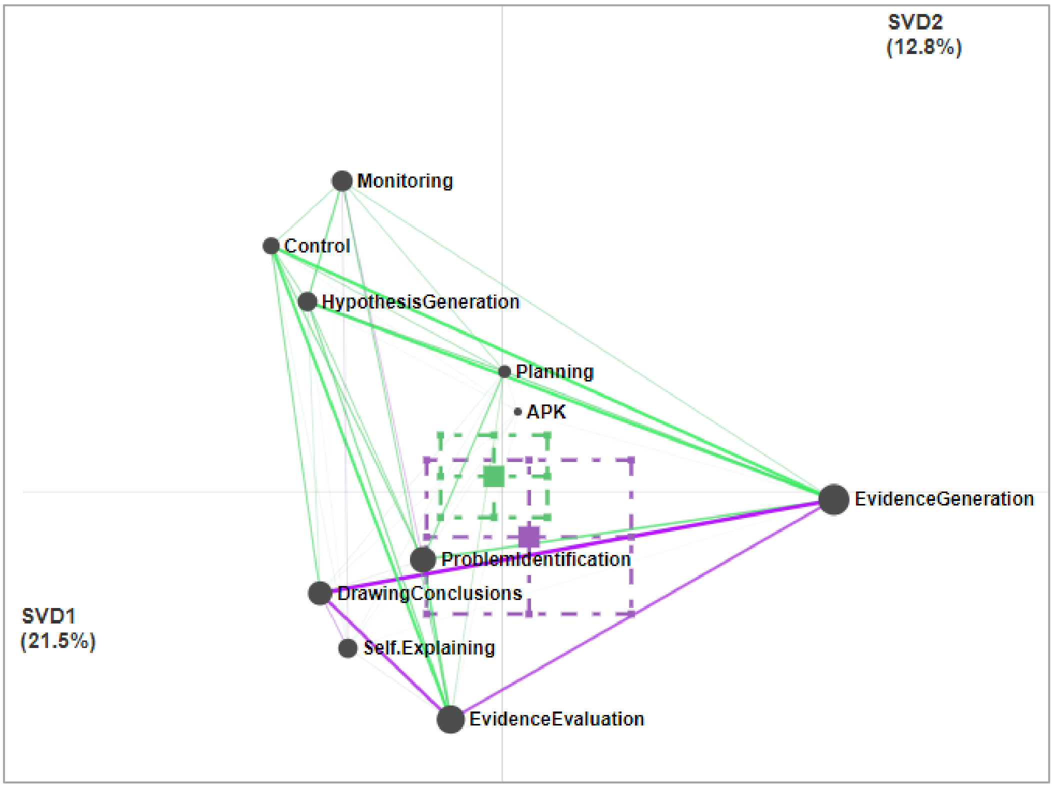
Transfer Task

Epistemic Network Analysis

VMP and VM vs. Control. Along the X axis, a two sample *t*-test assuming unequal variance showed that the position of the VMP and VM centroid (M = -0.04, SD = 0.89, N = 66) was not significantly different from the control (M = 0.11, SD = 0.94, N = 22; $t(34.29)$ = -0.63, p = .53, d = 0.16). Along the Y axis, the centroid of VMP and VM (M = 0.06, SD = 0.68, N = 66) was not significantly different from the control (M = -0.19, SD = 0.71, N = 22; $t(34.90)$ = -1.43, p = .16, d = 0.36). These findings indicate that there were no significant differences between the epistemic networks of the two video modeling conditions and the control condition in the transfer task. Figure B2 shows that participants in the video modeling conditions (green) made more connections between *control*, *planning*, *evidence generation* and *hypothesis generation*. Participants in the control condition (purple) were *monitoring* during *problem identification* and integrated *evidence generation*, *evidence evaluation*, and *drawing conclusions*. However, since we found no significant differences between the centroids, these findings are purely descriptive.

Figure B2

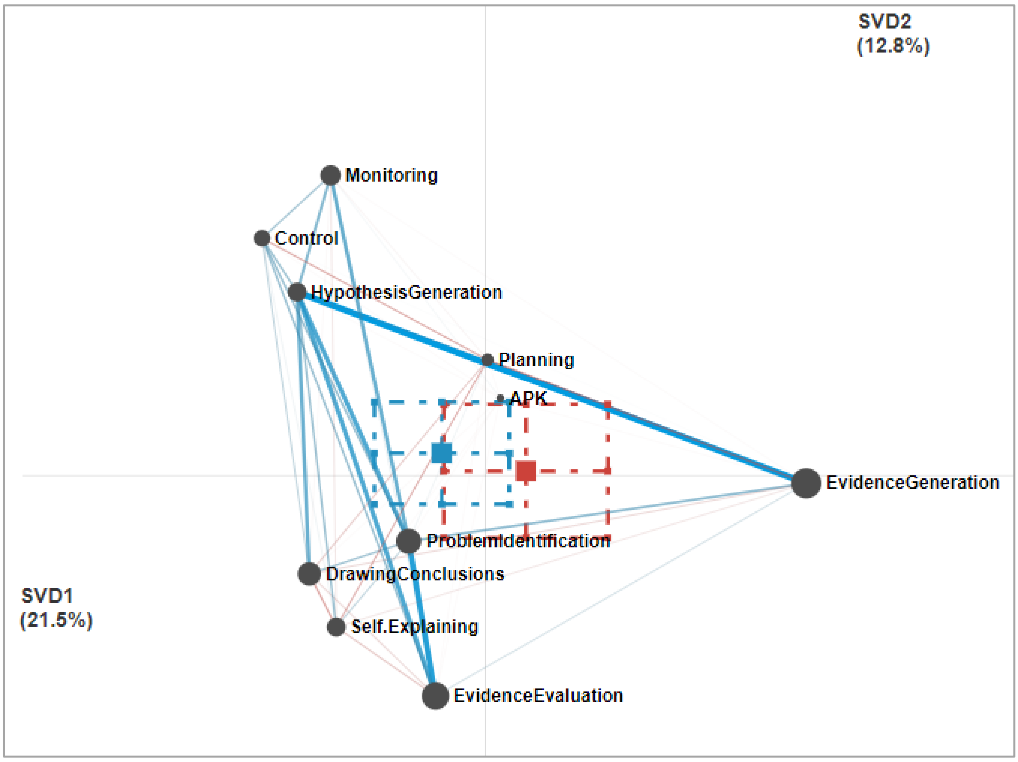
The Network Difference between the VMP and VM (green) and the Control (purple) Conditions



VMP vs. VM. Along the X axis, a two sample *t*-test showed that VMP (M = 0.17, SD = 0.89, N = 28) was not significantly different from VM (M = -0.19, SD = 0.86, N = 38; $t(57.37)$ = -1.62, p = .11, d = 0.40). Along the Y axis, the centroid of VMP (M = 0.02, SD = 0.73, N = 28) was not significantly different from VM (M = 0.09, SD = 0.66, N = 38; $t(54.76)$ = -0.43, p = .67, d = 0.11). In the transfer task, participants in the VM condition had stronger connections between *monitoring*, *control* and all other scientific reasoning processes than the VMP condition, see Figure B3.

Figure B3

The Network Difference between the VMP (red) and the VM (blue) Condition

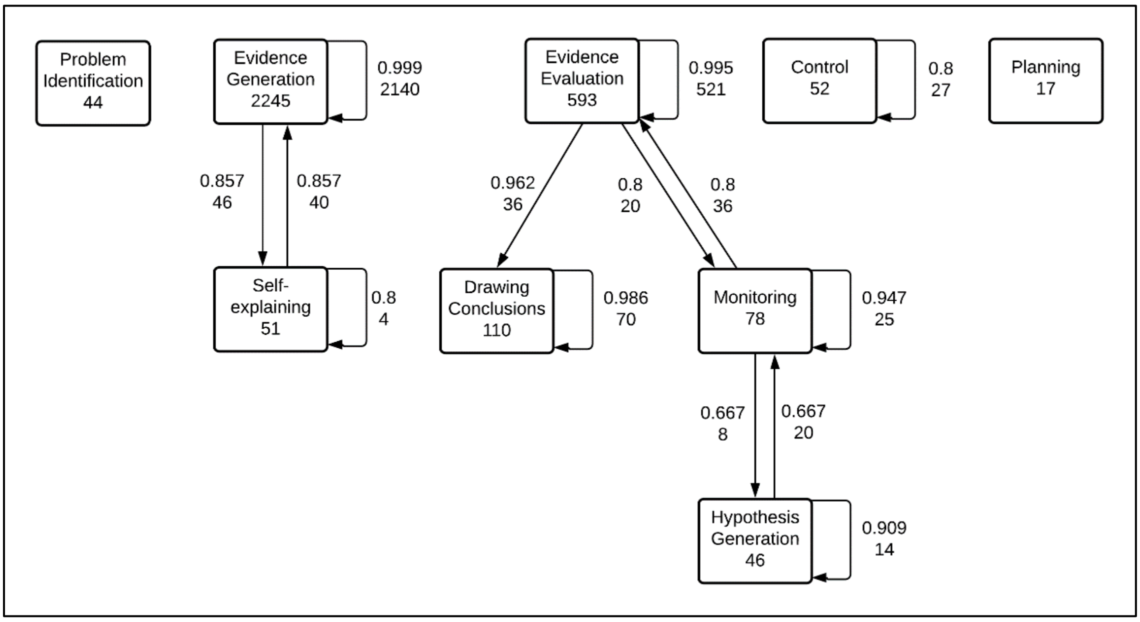


Process Mining

We used process mining to compare the video modeling conditions and the control condition. Figure B4 illustrates that Problem Identification, Control, and Planning were disconnected from other processes. Again, Control had self-loops, indicating that participants engaged in Control successively, but not consistently before or after another activity. There were strong reciprocal links between Evidence Generation and Self-Explaining, Hypothesis Generation and Monitoring, Monitoring and Evidence Evaluation. Participants were first evaluating evidence and then drew a conclusion, which corresponds to models of scientific reasoning.

Figure B4

The Process Model of the Two Video Modeling Conditions Combined (n = 66)

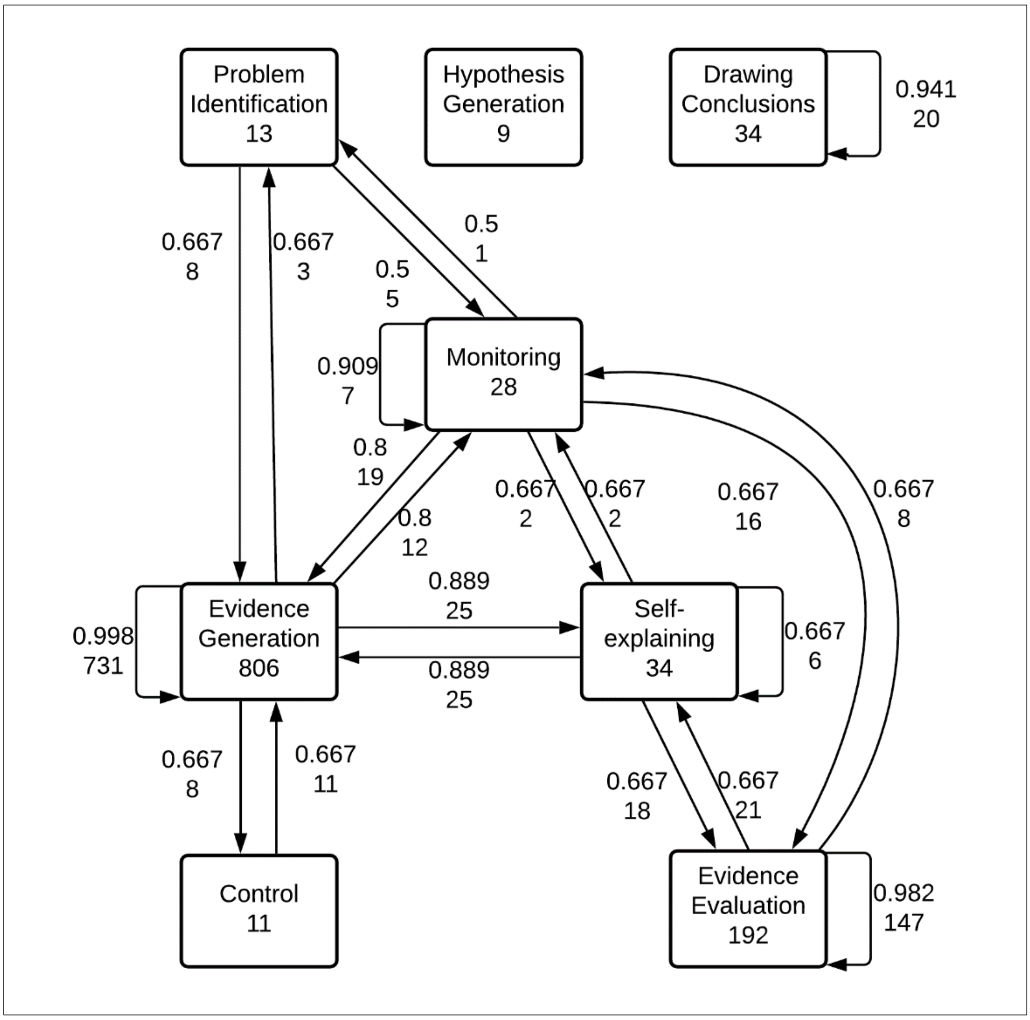


Note. The top number (referring to the arrows) represents the dependency (0-1), and the number below represents the frequency of each sequence. The numbers in the boxes represent the occurrence of each process.

In the control condition (Figure B5), Hypothesis Generation was disconnected from other processes like in the training task. Problem Identification was followed by Evidence Generation, however with a weaker dependency (0.67). Next, Evidence Generation had reciprocal relationships with Self-Explaining and Evidence Evaluation. Monitoring was connected to Problem Identification and Self-Explaining, but rather weakly and rarely (frequencies 6 and 4, respectively). Evidence Evaluation was the final activity in the model and Drawing Conclusions was disconnected from other processes.

Figure B5

The Process Model of the Control Condition (n = 22)



Note. 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.