Toward Lasting Learning?

Enhancing Non-Interactive Teaching in Inquiry-Based Authentic Science Lessons

Dissertation

zur Erlangung des Doktorgrades

der Wirtschafts- und Sozialwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

vorgelegt von

Heike Russ

Tübingen

2025

1. Betreuer:	Prof. Dr. Andreas Lachner
2. Betreuerin:	Prof. Dr. Katharina Scheiter

Tag der mündlichen Prüfung:	31.07.2025
Dekanin:	Prof. Dr. Taiga Brahm
Dekan:	Prof. Dr. Dominik Papies

1. Gutachter:	Prof. Dr. Andreas Lachner
2. Gutachter:	Prof. Dr. Thorsten Bohl

Danksagung

Mein erster Dank geht an die insgesamt 1.429 Schüler:innen der 60 Klassen, die in der 1,5-jährigen Erhebungszeit an meinem Physikprojekt teilgenommen haben. Danke an die gesamten Schulgemeinden, die dieses Projekt tatkräftig unterstützt und damit erst ermöglicht haben! Ich freue mich auf die weitere Zusammenarbeit.

Mein Dank gilt außerdem meinen beiden Supervisors, Prof. Dr. Andreas Lachner, der mich während des gesamten Projekts engagiert begleitet hat und meiner Arbeit stets vertraute, und Prof. Dr. Katharina Scheiter, die mich mit ihrem Feedback konstruktiv weiterbrachte.

Ich danke außerdem den weiteren Mitgliedern des Projektteams für die Unterstützung—Ass.-Prof. Leonie Sibley, Jun.-Prof. Salome Flegr, Prof. Dr. Jochen Kuhn, Ass.-Prof. Dr. Vincent Hoogerheide, Dr. Tino Endres (beteiligt an Studie 3)—wie auch der DFG Lasting Learning Forschungsgruppe, exemplarisch genannt seien hier Prof. Dr. Tobias Richter und Dr. Veit Kubik. Danke auch an meine drei "Hiwinen", die während der gesamten Projektzeit mit mir gearbeitet haben—Celine Alb, Katharina Groß und Anna Heinrich.

Für die vielen schönen und lustigen gemeinsamen Momente, wie auch die wichtigen Gespräche, die wir im "best room W31" hatten, bedanke ich mich bei meinem Schreibtischnebensitzer Marcel Capparozza, sowie bei Dr. Martí Quixal, Dr. Stephen Bodnar, Runa Marold und Eric Eichkorn. Mit euch würde ich immer wieder in einen Büroraum sitzen!

Ein ganz besonderer Dank gilt meiner Familie—meinem Mann Thomas sowie unseren Kindern Isabella, Nils und Charlotta, die immer Rücksicht genommen haben auf meine Termine, Videokonferenzen (😉) und sämtliche Übermüdungserscheinungen während der Erhebungsphasen. Vielen Dank auch an meine Eltern, die immer an mich glauben und mit zahlreichen Oma-Opa-Tagen unser Familiensystem aufrechterhalten haben.

Für die Übernahme der Begutachtungen bzw. des Prüfungsvorsitzes danke ich Prof. Dr. Andreas Lachner, Prof. Dr. Thorsten Bohl und Prof. Dr. Marcus Syring.

Summary

High-quality teaching is key to students' academic success but also essential for promoting educational equity. Particularly for students from disadvantaged backgrounds and socially deprived school contexts, classroom teaching often serves as the primary lever to reduce inequality. In this context, lasting subject-specific learning is a central criterion for successful teaching. However, recent large-scale achievement studies have highlighted that students often struggle to acquire and retain such lasting knowledge, especially in domains such as science. An effective way to enhance students' learning is to have them teach previously learned contents to others, a method known as *learning by teaching*. Recent research has shown that teaching contents even to a fictitious, non-present peer—referred to as *non-interactive teaching*—is also a powerful generative activity. It requires students to adapt their explanations to an imagined peer, who can be portrayed in various ways. The activity is also suitable for individual learning settings. Non-interactive teaching can facilitate the meaningful construction of coherent mental representations, contributing to learning and metacomprehension. However, this effect is modest and heterogeneous.

Therefore, we investigated how to optimize the generative activity of non-interactive teaching in three experimental classroom studies conducted in inquiry-based authentic physics lessons with secondary students ($N = 1{,}251$). We tested whether this activity can be enhanced through drawing, by distributing it throughout the study phase, or by sequentially adding retrieval practice. We assessed students' conceptual knowledge and monitoring accuracy regarding both immediate and lasting learning after eight weeks.

The results revealed that combining non-interactive teaching with drawing improved students' immediate conceptual knowledge, likely due to increased task interest and more complete explanations. In contrast, distributing the activity throughout the study phase did not yield additional benefits. No main effect emerged for the sequential combination of non-

interactive teaching and retrieval practice. However, exploratory analyses revealed that retrieval practice supported lasting learning when prior generative processing was of low quality—that is, when students' explanations were incomplete, less elaborated, or inaccurate.

Together, the findings demonstrate that non-interactive teaching can be systematically optimized to support learning in authentic science lessons. Carefully designed generative activities—combined with retrieval practice—offer a promising approach to fostering lasting learning in secondary science education. In this sense, the present dissertation contributes to strengthening successful teaching—and thus teaching quality—as a key lever for promoting students' academic success and educational equity. It provides practice-oriented insights into how evidence-informed teaching activities can be implemented in everyday classroom teaching. Moreover, it highlights how such approaches can support sustainable school development and underlines the importance of promoting high teaching quality across all phases of teacher education to improve student learning.

Zusammenfassung

Eine hohe Unterrichtsqualität ist entscheidend für den schulischen Erfolg von Schüler:innen und zugleich zentral für die Förderung von Bildungsgerechtigkeit. Besonders für Schüler:innen aus benachteiligten Herkunftskontexten und sozial deprivierten Schulumfeldern stellt der Unterricht häufig den wichtigsten Hebel dar, um Ungleichheiten auszugleichen. In diesem Zusammenhang ist dauerhaftes fachliches Lernen ein zentrales Kriterium erfolgreichen Unterrichts. Aktuelle internationale Leistungsstudien zeigen jedoch, dass viele Schüler:innen Schwierigkeiten haben, Wissen aufzubauen und dauerhaft zu behalten – insbesondere in den Naturwissenschaften.

Ein wirksamer Weg, um das Lernen von Schüler:innen zu fördern, besteht darin, sie zuvor Gelerntes anderen erklären zu lassen. Dieses Vorgehen ist als *Lernen durch Lehren* bekannt. Aktuelle Forschung zeigt, dass auch das Erklären gegenüber einer fiktiven, nicht anwesenden Person—bekannt als *nicht-interaktives Lehren*—eine wirksame generative Lernaktivität darstellt. Dabei müssen Schüler:innen ihre Erklärungen auf eine fiktive Person zuschneiden, die unterschiedlich dargestellt werden kann. Die Aktivität lässt sich zudem flexibel in individuellen Lernsituationen einsetzen. Nicht-interaktives Lehren kann zur bedeutungsvollen Konstruktion kohärenter mentaler Repräsentationen beitragen und somit Lernen und metakognitive Prozesse fördern. Die Wirksamkeit fällt jedoch insgesamt eher gering und auch uneinheitlich aus.

Vor diesem Hintergrund wurden im Rahmen dieser Dissertation drei experimentelle Unterrichtsstudien im Fach Physik mit insgesamt 1.251 Schüler:innen der Sekundarstufe durchgeführt. Ziel war es, die generative Aktivität des nicht-interaktiven Lehrens in authentischem Unterricht, der forschendes Lernen beinhaltet, zu optimieren. Es wurde untersucht, ob die Aktivität des nicht-interaktiven Lehrens durch die Kombination mit Zeichnen, durch das Verteilen der Lernaktivität über die Studierphase oder durch die

Kombination mit Retrieval Practice (Abrufübung) verbessert werden kann. Erfasst wurden das konzeptuelle Wissen und die Genauigkeit der Wissenseinschätzung der Schüler:innen—sowohl unmittelbar als auch mit einer zeitlich verzögerten Erhebung nach acht Wochen.

Die Ergebnisse zeigten, dass die Kombination des nicht-interaktiven Lehrens mit Zeichnen kurzfristige Lerngewinne förderte. Die Effekte erklärten sich durch gesteigertes aufgabenbezogenes Interesse und vollständigere Erklärungen der Schüler:innen. Die mehrfache Durchführung der Lernaktivität des nicht-interaktiven Lehrens führte hingegen zu keinen zusätzlichen Effekten. Auch die Kombination mit Retrieval Practice zeigte keine Haupteffekte. In explorativen Analysen zeigte sich jedoch, dass Retrieval Practice dann zu langfristigem Lernen beitrug, wenn die vorherigen Erklärungen der Lernenden von geringer Qualität waren—also unvollständig, wenig elaboriert oder fehlerhaft.

Insgesamt verdeutlichen die Ergebnisse, dass nicht-interaktives Lehren systematisch optimiert werden kann, um das Lernen im regulären naturwissenschaftlichen Unterricht wirksam zu unterstützen. Sorgfältig konzipierte generative Lernaktivitäten—kombiniert mit Retrieval Practice—stellen einen vielversprechenden Ansatz dar, um nachhaltiges Lernen im Fachunterricht der Sekundarstufe zu fördern. Die Dissertation leistet damit einen Beitrag zur Stärkung erfolgreichen Unterrichts—und damit der Unterrichtsqualität—als zentralem Hebel für schulischen Erfolg und Bildungsgerechtigkeit. Sie liefert praxisnahe Einblicke, wie evidenzbasierte Lernaktivitäten im Schulalltag umgesetzt werden können, zeigt Anknüpfungspunkte für nachhaltige Schulentwicklungsprozesse auf und betont die Bedeutung hoher Unterrichtsqualität in allen Phasen der Lehrerbildung zur Förderung des Lernens aller Schüler:innen.

List of Publications and Declaration of Author's Contribution

The following list contains all publications included in this dissertation.

Publication 1 (content of Study 1) has been accepted for publication in the *Journal of Educational Psychology* and is currently in press.

Publication 2 (content of Study 2) has been accepted for publication and has been published in a special issue of *Learning and Individual Differences*.

Publication 3 (content of Study 3) is currently under review. It is intended for inclusion in a special issue of a peer-reviewed journal.

The proportional contribution of all authors is listed below.

The three publications are presented in Chapters 3 to 5.

Publication of Study 1

Author	Author position	Scientific ideas	Data generation	Analysis & Interpretation	Paper writing
Heike Russ	First author	30%	100%	85%	75%
Leonie Sibley	Second author			5%	5%
Salome Flegr	Third author	5%			
Jochen Kuhn	Fourth author	5%			
Vincent Hoogerheide	Fifth author				5%
Katharina Scheiter	Sixth author	30%			5%
Andreas Lachner	Seventh author	30%		10%	10%

Title of paper:	Combining Non-Interactive Teaching and Drawing Fosters Conceptual Knowledge but not Monitoring Accuracy From Guided Inquiry in Science Learning
Status:	Accepted in the *Journal of Educational Psychology* (open access): Russ, H., Sibley, L., Flegr, S., Kuhn, J., Hoogerheide, V., Scheiter, K., & Lachner, A. (in press). Combining non-interactive teaching and drawing fosters conceptual knowledge but not monitoring accuracy from guided inquiry in science learning. *Journal of Educational Psychology*. https://doi.org/10.1037/edu0000971
Open Science Badges:	Preregistered+ TC (preregistered including analysis plan; transparent changes), Open Materials, Open Data, Open Code

Publication of Study 2

Author	Author position	Scientific ideas	Data generation	Analysis & Interpretation	Paper writing
Heike Russ	First author	30%	100%	85%	80%
Leonie Sibley	Second author			5%	5%
Salome Flegr	Third author	5%			
Jochen Kuhn	Fourth author	5%			
Vincent Hoogerheide	Fifth author				5%
Katharina Scheiter	Sixth author	30%			
Andreas Lachner	Seventh author	30%		10%	10%

Title of paper:	Does Distributing Non-Interactive Teaching Contribute to Learning? Students' Academic Self-Concept and Work Ethic Matter
Status:	Published in Learning and Individual Differences (open access): Russ, H., Sibley, L., Flegr, S., Kuhn, J., Hoogerheide, V., Scheiter, K., & Lachner, A. (2025). Does distributing non-interactive teaching contribute to learning? Students' academic self-concept and work ethic matter. *Learning and Individual Differences,* 120, 102687. https://doi.org/10.1016/j.lindif.2025.102687 [This article is part of a special issue entitled: 'Do desirable difficulties benefit all? On the role of individual differences for the effects of desirable difficulties in meaningful learning' published in *Learning and Individual Differences*.]
Open Science Badges:	Preregistered+ (preregistered including analysis plan), Open Materials, Open Data, Open Code

Publication of Study 3

Author	Author position	Scientific ideas	Data generation	Analysis & Interpretation	Paper writing
Heike Russ	First author	25%	100%	85%	85%
Tino Endres	Second author	5%			5%
Leonie Sibley	Third author	5%			
Salome Flegr	Fourth author	5%			
Jochen Kuhn	Fifth author	5%			
Vincent Hoogerheide	Sixth author	5%			5%
Katharina Scheiter	Seventh author	25%			
Andreas Lachner	Eighth author	25%		15%	5%

Title of paper:	Happy or Redundant? Retrieval Practice Enhances Lasting Learning in Non-Interactive Teaching for Low-Quality Explanations
Status:	Under review in a peer-reviewed journal: Russ, H., Endres, T., Sibley, L., Flegr, S., Kuhn, J., Hoogerheide, V., Scheiter, K., & Lachner, A. (under review). Happy or redundant? Retrieval practice enhances lasting learning in non-interactive teaching for low-quality explanations. [This article is intended for inclusion in a special issue of a peer-reviewed journal.]
Open Science Badges:	Preregistered+ TC (preregistered including analysis plan; transparent changes), Open Materials, Open Data, Open Code

Declaration of generative AI and AI-assisted technologies in the writing process

The following statement is included in each of the journal articles and is also valid for the framing parts of this cumulative dissertation authored solely by the doctoral candidate:

During the preparation of this work the author(s) used ChatGPT-4o in order to ensure error-free spelling and to receive suggestions for more suitable phrasing. After using this tool, the author(s) reviewed and edited the content as needed and take full responsibility for the content of the publication.

Table of Contents

1 Introduction and Theoretical Background 3

1.1 Teaching and Learning 5

1.2 Generative Learning 9

1.3 Learning by Non-Interactive Teaching 10

1.4 Enhancing the Effectiveness of Non-Interactive Teaching 11

1.4.1 Enhancing Non-Interactive Teaching With Drawing 11

1.4.2 Enhancing Non-Interactive Teaching Through Distribution 12

1.4.3 Enhancing Non-Interactive Teaching With Retrieval Practice 14

1.5 Situating Non-Interactive Teaching Within Contexts of Teaching and Learning 15

1.5.1 Non-Interactive Teaching Within Teaching and Learning 15

1.5.2 Non-Interactive Teaching Within the Offer-Use Model of Teaching 16

2 The Dissertation's Research Focus 21

2.1 Aims and Research Questions 21

2.2 Locating the Research Focus Within the Offer-Use Model of Teaching for Students' Non-Interactive Teaching 24

3 Study 1: Combining Non-Interactive Teaching and Drawing Fosters Conceptual Knowledge but not Monitoring Accuracy From Guided Inquiry in Science Learning ... 29

3.1 Introduction 30

3.1.1 Enhancing Conceptual Understanding in Science Education Through Inquiry Learning 31

3.2 Generative Activities to Foster Inquiry-Based Learning 33

3.2.1 Learning by Non-Interactive Teaching 34

3.2.2 Learning by Drawing 37

3.2.3 Combining Non-Interactive Teaching and Drawing 39

3.3 The Present Study ... 40
3.3.1 Generation Hypothesis ... 41
3.3.2 Visualization Hypothesis ... 41
3.3.3 Drawing Hypothesis ... 42
3.3.4 Further Explorative Analyses ... 42
3.4 Method ... 42
3.4.1 Participants and Design ... 42
3.4.2 The Teaching Unit ... 44
3.4.2.1 Introduction to the Converging Lens ... 44
3.4.2.2 Students' Experiments ... 44
3.4.3 Learning Activity ... 46
3.4.4 Measures ... 49
3.4.4.1 Cognitive Learning Outcomes ... 49
3.4.4.2 Metacognitive Learning Outcomes ... 50
3.4.4.3 Task-Specific Motivation ... 51
3.4.4.4 Additional Control Measures ... 51
3.4.5 Procedure ... 52
3.4.6 Analysis and Coding of Students' Explanations and Drawings ... 55
3.4.6.1 Characteristics of Students' Explanations ... 55
3.4.6.2 Characteristics of Students' Drawings ... 56
3.4.7 Data Analyses ... 57
3.4.8 Transparency and Openness ... 58
3.5 Results ... 58
3.5.1 Preliminary Analysis ... 58
3.5.2 Cognitive Learning Outcomes ... 62
3.5.3 Metacognitive Learning Outcomes ... 68

3.6 Discussion 68

3.6.1 Limitations and Future Directions 75

3.7 Conclusion 80

4 Study 2: Does Distributing Non-Interactive Teaching Contribute to Learning? Students' Academic Self-Concept and Work Ethic Matter 83

4.1 Introduction 84

4.2 Learning by Non-Interactive Teaching 84

4.2.1 Enhancing the Effectiveness of Non-Interactive Teaching 86

4.2.1.1 Enhancing Non-Interactive Teaching With Drawing 87

4.2.1.2 Enhancing Non-Interactive Teaching Through Distribution 88

4.2.2 Influence of Students' Individual Differences 90

4.3 The Present Study 93

4.3.1 Generation Hypothesis (H1) 93

4.3.2 Drawing Hypothesis (H2) 93

4.3.3 Distribution Hypothesis (H3) 94

4.3.4 Explorative Analyses on Students' Individual Differences 94

4.4 Method 94

4.4.1 Participants and Design 94

4.4.2 The Teaching Unit 95

4.4.2.1 Introduction to the Converging Lens 96

4.4.2.2 Students' Experiments 96

4.4.3 Learning Activity 97

4.4.4 Measures 99

4.4.4.1 Conceptual Knowledge 100

4.4.4.2 Monitoring Accuracy 100

4.4.4.3 Academic Self-Concept in Physics 101

4.4.4.4 Physics Work Ethic 101

4.4.5 Procedure 101

4.4.6 Analysis of Students' Experimentation Worksheets 104

4.4.7 Analysis of Students' Explanations, Drawings, and Restudy Notes 104

4.4.7.1 Completeness of Students' Explanations, Drawings, and Restudy Notes 105

4.4.7.2 Elaboration of Students' Explanations, Drawings, and Restudy Notes 105

4.4.7.3 Correctness of Students' Explanations, Drawings, and Restudy Notes 106

4.4.8 Data Analyses 106

4.5 Results 108

4.5.1 Preliminary Analysis 108

4.5.2 Conceptual Knowledge 111

4.5.3 Monitoring Accuracy 112

4.5.4 Explorative Analyses on Students' Individual Differences 113

4.5.4.1 Prior Knowledge 113

4.5.4.2 Academic Self-Concept 113

4.5.4.3 Physics work ethic 115

4.6 Discussion 116

4.6.1 Study Limitations and Future Research 119

4.7 Conclusion 121

5 Study 3: Happy or Redundant? Retrieval Practice Enhances Lasting Learning in Non-Interactive Teaching for Low-Quality Explanations 125

5.1 Introduction 126

5.2 Generative Learning Activities 127

5.3 Retrieval Practice 128

5.4 Combining Generative Activities and Retrieval Practice 129

5.5 The Present Study 132

5.5.1 Generation Hypothesis (H1) 132

5.5.2 Retrieval Hypotheses (H2, H3) 132

5.5.3 Interaction Hypothesis (H4) 133

5.5.4 Explorative Analyses on the Characteristics of Students' Explanations 133

5.6 Method 133

5.6.1 Participants and Design 133

5.6.2 The Teaching Unit 134

5.6.3 Generative Learning Activity 135

5.6.4 Retrieval Practice 137

5.6.5 Measures 139

5.6.5.1 Conceptual Knowledge 139

5.6.5.2 Monitoring Accuracy 140

5.6.5.3 Additional Control Measures 140

5.6.6 Procedure 141

5.6.7 Analysis of Students' Explanations, Drawings, and Restudy Notes 143

5.6.8 Analysis of Students' Retrieval Quiz 144

5.6.9 Data Analyses 144

5.7 Results 145

5.7.1 Preliminary Analysis 145

5.7.2 Does Engaging in a Generative Learning Activity Enhance Immediate and Lasting Learning? 146

5.7.3 Does Retrieval Practice Enhance Immediate and Lasting Learning? 146

5.7.4 Does the Sequential Combination of a Generative Learning Activity and Retrieval Practice Yield Synergistic Effects on Immediate and Lasting Learning? 146

5.7.5 Does the Effectiveness of Retrieval Practice Depend on the Quality of Students' Generative Processing? 147

5.8 Discussion 149

5.8.1 Study Limitations and Future Research 151

5.9 Conclusion.. 152

6 General Discussion 157

6.1 Discussion of the Results 157

6.1.1 Summary of the Results 157

6.1.2 Situating the Results Within Prior Research and the Offer-Use Model of Teaching for Students' Non-Interactive Teaching 159

6.1.3 Implications for the Offer–Use Model of Teaching for Students' Non-Interactive Teaching 165

6.2 Implications for Educational Research 170

6.2.1 Relevance and Placement Within the Field 170

6.2.2 Methodological Reflections on Research in Authentic Classroom Settings 171

6.2.3 Deliberate Research Focus and Analytical Scope 172

6.3 Implications for Educational Practice 173

6.3.1 Implications for Classroom Teaching.. 173

6.3.2 Implications for School Development.. 175

6.3.3 Implications for Teacher Education... 176

6.4 Limitations and Directions for Future Research 178

6.5 Conclusion.. 180

7 References 185

8 Appendix 219

8.1 Appendix Study 1 219

8.2 Appendix Study 2 224

8.3 Appendix Study 3 234

List of Figures

Figure 1. Offer-Use Model of Teaching 8

Figure 2. The Select-Organize-Integrate (SOI) Model of Generative Learning 10

Figure 3. Offer-Use Model of Teaching for Students' Non-Interactive Teaching 18

Figure 4. The Dissertation's Research Focus located in the Offer-Use Model of Teaching for Students' Non-Interactive Teaching 26

Figure 5. Experimental Setup on the Converging Lens and its Images 45

Figure 6. Mockup Messenger Chat fpr Generative Conditions 48

Figure 7. Exemplary Item From the ROC-CI Conceptual Knowledge Test About Converging Lenses 50

Figure 8. Design and Procedure of the Study 54

Figure 9. Mediation Analysis in Terms of Task Interest and Task Enjoyment 64

Figure 10. Serial Mediation Analysis in Terms of Task Interest and Completeness of Students' Explanations 66

Figure 11. Students' Experimental Setup on the Converging Lens and its Images 97

Figure 12. Mock-up Messenger Chat for the Generative Conditions (Teaching-Only, Teaching + Drawing) 99

Figure 13. Explorative Mediation Analysis in Terms of the Level of Completeness 111

Figure 14. Johnson-Neyman Confidence Intervals for Academic Self-Concept Moderation of Immediate Conceptual Knowledge in After-Study and Distributed Conditions 114

Figure 15. Johnson-Neyman Confidence Intervals for Physics Work Ethic Moderation of Immediate Conceptual Knowledge in After-Study and Distributed Conditions 116

Figure 16. Setup of the Students' Experiments on the Converging Lens and its Images 135

Figure 17. Mockup Messenger Chat of the Generative Learning Activity 136

Figure 18. Example Question From the Retrieval Quiz 138

Figure 19. Sample Item From the ROC-CI Conceptual Knowledge Test on Converging Lenses 139

Figure 20. Design and Procedure of the Study 143

Figure 21. Moderation of Completeness, Elaboration, and Correctness on Lasting Conceptual Knowledge in Generative and Non-Generative Conditions 148

Figure 22. Offer-Use Model of Teaching for Students' Non-Interactive Teaching With Results From Study 1, 2, and 3 164

Figure 23. Expanded Offer-Use Model of Teaching for Students' Non-Interactive Teaching 169

List of Tables

Table 1. Means and Standard Deviations for all Measurements Across Experimental Conditions 60

Table 2. Summary of the Contrast Analyses on (Meta-) Cognitive Learning Outcomes 65

Table 3. Summary of the Mediation Analyses With Task Interest and Characteristics of Students' Explanations Regarding Immediate Cognitive Learning Outcome 67

Table 4. Procedure of the Study 102

Table 5. Means and Standard Deviations for all Measurements Across Experimental Conditions 110

1

INTRODUCTION AND THEORETICAL BACKGROUND

1 Introduction and Theoretical Background

High-quality teaching is widely regarded as one of the most influential school-related factors affecting student learning and achievement (Bohl, 2023; Klieme, 2022; Lipowsky, 2020). Its importance extends beyond general educational outcomes: successful teaching plays a particularly critical role in supporting students from disadvantaged backgrounds and those attending schools in socially deprived areas (Bremm et al., 2016; Klein, 2017). For these students, classroom teaching is often the most powerful lever to compensate for unequal starting conditions and limited learning opportunities outside of school (Baumert et al., 2010; Cadima et al., 2010; Decristan et al., 2016; Lipowsky & Bleck, 2019; Seiz et al., 2016). Accordingly, ensuring teaching quality is not only a question of effectiveness but also of educational equity.

In this context, one key criterion of successful teaching is students' subject-specific learning (Klieme, 2022), which should ideally be lasting rather than temporary (see T. Richter et al., 2022). Indeed, stable knowledge helps students to make sense of new information, critically assess it, and reach informed decisions (T. Richter et al., 2022). Lasting knowledge is therefore essential for students' academic success and their participation in today's knowledge society (OECD, 2016). Accordingly, school teachers carry great responsibility in supporting students to acquire and retain knowledge as a central educational goal. However, recent large-scale achievement studies (e.g., Mullis et al., 2020; OECD, 2023) indicated that students often struggle to acquire such lasting knowledge—particularly in domains such as science. Even with targeted approaches intended to foster scientific literacy, such as inquiry-based learning where students mimic the professional activities of researchers under supervised conditions (Pedaste et al., 2015), students tend to process contents superficially (Goldwater & Schalk, 2016; Schneider et al., 2011). This may be due to the high cognitive and metacognitive demands associated with such an instructional method (Alfieri et al., 2011).

Based on generative learning theory, it can be assumed that the generative learning activity of non-interactive teaching[1] can help facilitate the meaningful construction of coherent mental representations in inquiry-based settings (Fiorella, 2023b). During non-interactive teaching, students verbally explain contents to a fictitious, non-present peer, which can foster both students' conceptual knowledge and monitoring accuracy (Lachner et al., 2022), that is, the accuracy of judgments about one's own understanding (Schraw, 2009). The approach of non-interactive teaching could therefore be a promising way to enable successful teaching in subject-specific contexts (see also Klieme, 2022).

However, the overall effect of non-interactive teaching is relatively modest (Kobayashi, 2024; Lachner et al., 2021) and heterogeneous (e.g., Hoogerheide, Visee, et al., 2019; Jacob et al., 2022). These findings suggest that non-interactive teaching alone is not necessarily effective and may require modifications. This raises the question of how the generative activity of non-interactive teaching regarding students' conceptual knowledge and monitoring accuracy can be optimized, particularly in the context of inquiry-based authentic science lessons in school and with respect to both immediate and lasting learning. Accordingly, and following recent research efforts targeting instructional improvement (see, e.g., Fiorella & Kuhlmann, 2020; Lachner et al., 2020; Pan et al., 2024; Roelle, Endres, et al., 2023), the present dissertation aims to address this question by systematically investigating complementary instructional modifications of non-interactive teaching across three empirical classroom studies. Based on a proposed Offer-Use Model of Teaching for Students' Non-Interactive Teaching (adapted from Vieluf et al., 2020; Fiorella & Mayer, 2016), each study focuses on a specific enhancement that may improve the effectiveness of non-interactive teaching in inquiry-based authentic science education. As

[1] The term *non-interactive teaching* refers to a form of *learning by teaching* in which students generate instructional explanations directed at an imaginary audience. This includes adapting the explanation to the assumed knowledge and needs of the addressee (Lachner et al., 2022). It is therefore conceptually distinct from *learning by explaining*, which also encompasses forms such as self-explaining.
Importantly, the term *teaching* in *non-interactive teaching* refers to the explanatory activity performed by students and should not be confused with the broader notion of *teaching* used throughout this dissertation to denote formal classroom teaching.

potential enhancements of non-interactive teaching, Study 1 investigates the combination of non-interactive teaching and drawing, Study 2 the distribution of non-interactive teaching across multiple points of a study phase, and Study 3 the combination of non-interactive teaching and retrieval practice. All three studies investigate specific instructional modifications aimed at improving students' conceptual knowledge and monitoring accuracy in inquiry-based authentic science lessons, with regard to both immediate and lasting learning. Together, these studies aim to provide a systematic, theoretically and empirically grounded, and practically relevant contribution to optimizing non-interactive teaching as a means to foster learning in real-world science education in school. In doing so, these three studies also aim to contribute to enabling successful teaching in subject-specific contexts.

The present dissertation is structured as follows: Chapter 1 outlines the relevance of enhancing the generative learning activity of non-interactive teaching and introduces the theoretical background. Chapter 2 specifies the dissertation's research focus and the related contributions of the three empirical classroom studies. Chapters 3 to 5 present these studies: Chapter 3 is about enhancing non-interactive teaching with drawing, Chapter 4 through its distribution across a study phase, and Chapter 5 through its combination with retrieval practice. Finally, Chapter 6 discusses the findings across studies, their implications for educational research and practice, and outlines directions for future research.

1.1 Teaching and Learning

According to Terhart (2019), the term "teaching" refers to situations that are characterized by four essential features: (1) a pedagogical intention, (2) a systematic and planned organization, (3) an institutional setting[2], and (4) the professional activity of teachers aimed at enhancing students' knowledge and skills. Against this background and with regard to

[2] In line with Terhart (2019), the term "institutional setting" refers specifically to the instructional context of general education in the formal school system. Other forms of instruction—such as in private music or language schools—are not considered within the scope of this definition.

teaching quality, successful teaching is characterized by the achievement of its educational objectives (Berliner, 2005, for a normatively oriented understanding of teaching quality, see e.g., Berliner, 2005; Klafki, 1995). These outcomes of teaching may include knowledge, understanding, skills, and competencies, the development of students' self-concept, domain-specific interest, and motivation, as well as psychosocial and emotional aspects (Berliner, 2005; Klieme, 2022). From an educational science perspective, the focus is not solely on output measurement; rather, the teaching process itself is central and examined systematically in terms of its effects on student learning (Klieme, 2022). Successful subject-specific teaching is particularly important for students from disadvantaged backgrounds or those attending schools in socially deprived contexts (Bremm et al., 2016; Klein, 2017). For these students, for example, teaching quality and student background factors are particularly closely intertwined, and high-quality teaching can play a key role in supporting their learning and compensating for unequal starting conditions (e.g., Cadima et al., 2010; Decristan et al., 2016; Seiz et al., 2016).

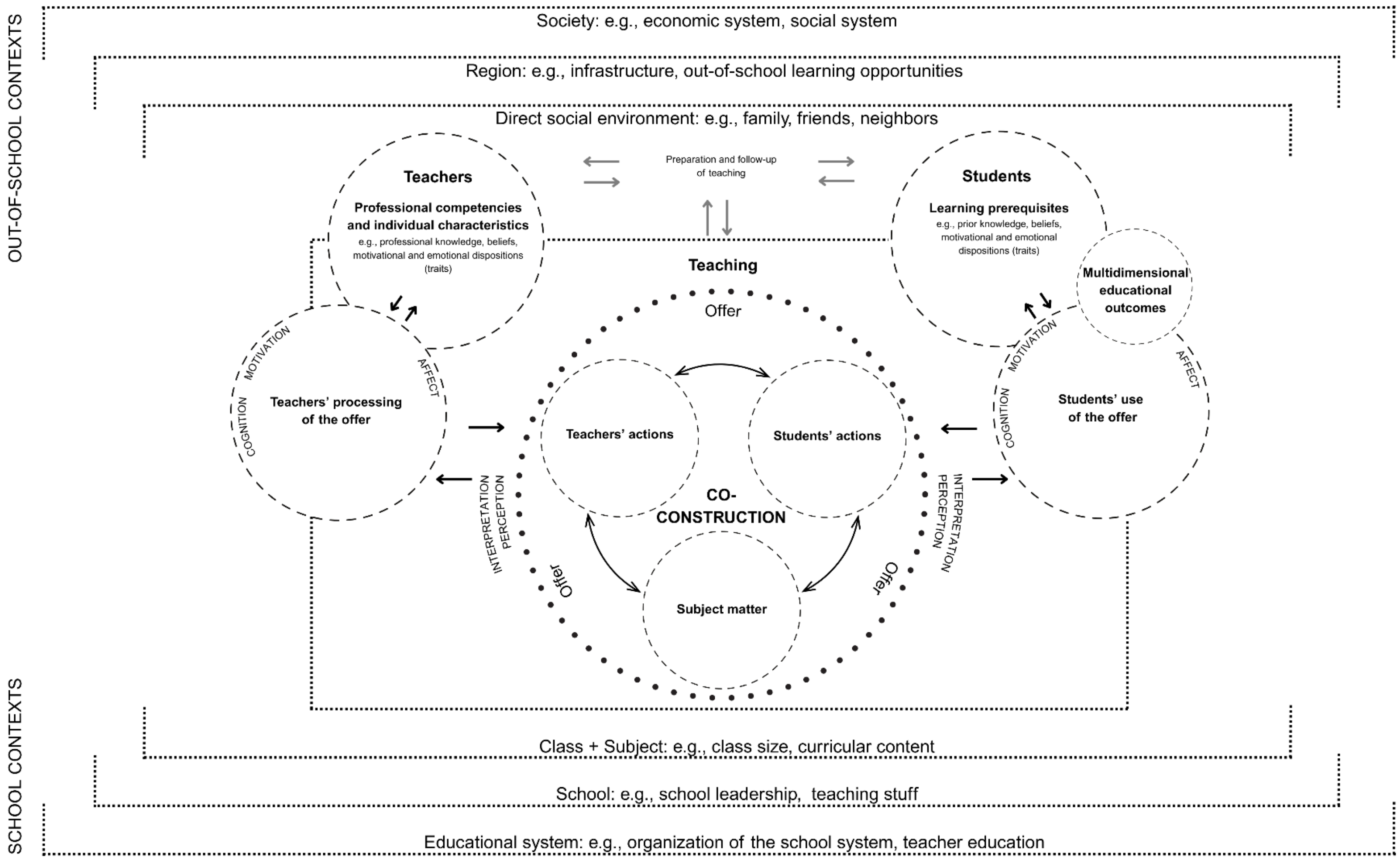
Overall, and following the current discourse on teaching (e.g., Keller et al., 2025; Klieme, 2022; Vieluf et al., 2020), teaching can be understood as an interactive process involving teachers, students, and the subject matter. Moreover, teaching can be understood as an offer that students—also taking into account their individual learning prerequisites—can use as co-constructors of their own learning process. This perspective is summarized in the currently prominent Offer-Use Model of Teaching (Figure 1) proposed by Vieluf et al. (2020). This model focuses on the reciprocal relationship between teaching offer and student use. It emphasizes that both teachers and students possess individual characteristics—such as prior knowledge, beliefs, and motivation—that shape how the teaching offer is processed and used. By embedding classroom interactions in multilayered school and out-of-school contexts and highlighting reciprocal influences, the model captures the complexity of teaching (in contrast to more linear models, see Fend, 2008) and accommodates the potential for diverse educational outcomes.

While the model of Vieluf et al. (2020) integrates both the teaching offer and its use in classroom settings, connecting the perspective of teaching and the perspective of learning is not straightforward, as teaching and learning, despite being interrelated, are not simply two sides of the same process. In contrast to teaching, the concept of learning typically centers on the individual cognitive processes of students.

Learning can be approached from different disciplinary perspectives. From a pedagogical perspective, learning is about the modalities of how learning takes place, but also about its content—namely, what learning does to the student and the world (Göhlich et al., 2014), and constructivist educational perspectives emphasize that knowledge is actively constructed through individual experience, prior knowledge, and interpretive processes, including within teaching contexts (Gudjons & Traub, 2020). According to the traditional, cognitively oriented research on teaching and learning, learning refers to the process of changing a student's knowledge or behavior, encompassing various forms of actions and behaviors. Typically, learning involves the acquisition of new knowledge and the reorganization of existing knowledge structures (Roelle, Lachner, et al., 2023). In this context, knowledge is defined as information that is stored relatively stably in long-term memory (Sweller, 2024). In order to successfully acquire new information, it is essential that students already possess relevant prior knowledge in long-term memory (Renkl, 2009). Such prior knowledge enables them to interpret new information meaningfully, assess conflicting perspectives, and construct coherent mental models—mental representations necessary for solving real-world problems and for meaningful interaction with their environment (Ifenthaler & Seel, 2011; Johnson-Laird, 1989; T. Richter & Maier, 2017). Therefore, fostering lasting learning is particularly important in school contexts. Lasting learning refers to durable changes in students' knowledge and cognitive skills that persist over time (T. Richter et al., 2022). Beyond short-term performance, an important goal of teaching is to help students build robust knowledge structures that they can access and flexibly apply across different situations and over extended periods of time.

Figure 1

Offer-Use Model of Teaching



Note. Adapted from Vieluf et al. (2020). Translated from German.

1.2 Generative Learning

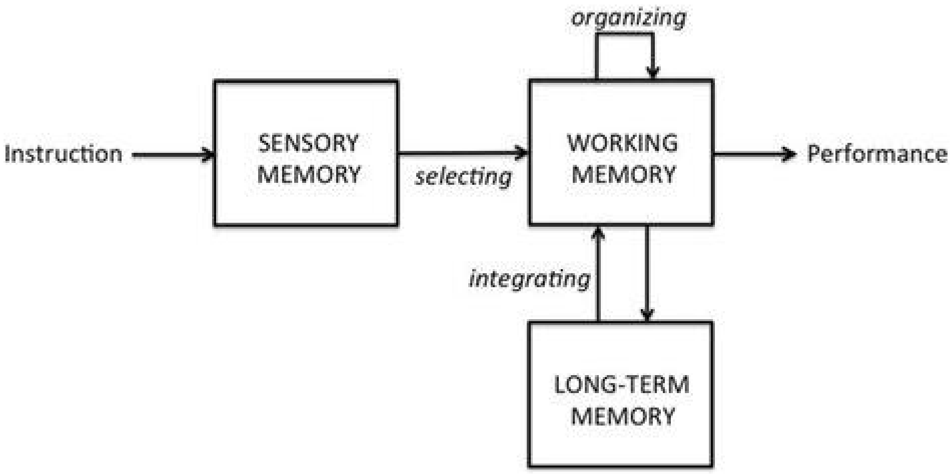
A central practical concern arising from the goal of fostering (lasting) learning in classrooms—and thereby enabling successful subject-specific teaching—is how students can be effectively supported in actively building robust knowledge. Generative learning offers a well-established and classroom-relevant theoretical approach to address this challenge.

Grounded in Wittrock's generative model of learning (1989, 2010), generative activities enhance the active construction of knowledge and are regarded as central to fostering meaningful learning in educational contexts (e.g., Fiorella, 2023b). One of the most widely recognized models of generative learning is the select-organize-integrate (SOI) model (Fiorella & Mayer, 2016), which outlines specific cognitive processes and primary memory stores for meaningful learning (see Figure 2). According to the SOI model, the generative learning process unfolds in three stages: First, students must select relevant information from the learning contents. Second, they need to organize this information in working memory in order to construct a coherent mental representation. Third, students must integrate this new mental representation with prior knowledge stored in long-term memory.

In line with the SOI model (Fiorella & Mayer, 2016), metacognitive and motivational processes are also important for successful generative learning. Metacognition involves knowing which information to select, which knowledge structure to construct, and which prior knowledge to activate during learning. Thus, learners' ability to accurately evaluate their own understanding of the contents is essential for engaging effectively in generative learning. In addition, motivation is central to initiating and sustaining generative processing, as it can be understood as a cognitive state that triggers, stimulates, and maintains goal-directed behavior throughout the learning process (Fiorella & Mayer, 2016).

Figure 2

The Select-Organize-Integrate (SOI) Model of Generative Learning



Note. From Fiorella & Mayer (2016).

1.3 Learning by Non-Interactive Teaching

One effective generative learning activity shown to enhance student learning is non-interactive teaching, in which students provide a verbal instructional explanation of previously learned contents to a fictitious, non-present peer (Lachner et al., 2022). During teaching, students engage in generative processes, which should yield better learning and metacomprehension (Fiorella, 2023b; Jacob et al., 2020).

Recent meta-analyses have shown small positive effects of non-interactive teaching on students' learning (Kobayashi, 2024: g = 0.27; Lachner et al., 2021: g = 0.22 for conceptual knowledge, g = 0.16 for transfer; Ribosa & Duran, 2022: g = 0.17) and some evidence has indicated benefits for students' monitoring accuracy. However, most studies relied on laboratory settings with university students and immediate posttests, and when delayed assessments were included, they were predominantly administered after one week (e.g., Fiorella & Mayer, 2013, 2014; Hoogerheide et al., 2014, 2016). Thus, it remains unclear whether non-interactive teaching is similarly effective in authentic classroom settings in schools (see also Sibley et al., 2024) and whether its benefits persist over time.

1.4 Enhancing the Effectiveness of Non-Interactive Teaching

Although the potential benefits of non-interactive teaching have been documented by recent meta-analyses (Kobayashi, 2024; Lachner et al., 2021; Ribosa & Duran, 2022), the reported findings vary considerably both between and within studies, with effects ranging from positive to null or even negative. This variability indicates that non-interactive teaching, in its current form, may not consistently lead to improved learning outcomes and points to the need for modification.

To address this, researchers have recently begun to explore three primary strategies for enhancing the effectiveness of non-interactive teaching. First, they have combined non-interactive teaching with additional generative learning activities such as drawing (e.g., Fiorella & Kuhlmann, 2020). Second, they have investigated ways to optimize the timing of non-interactive teaching (i.e., distribution, see Lachner et al., 2020), and third, they have examined how to combine generative activities with retrieval practice (Roelle, Endres, et al., 2023).

1.4.1 Enhancing Non-Interactive Teaching With Drawing

One promising approach to enhance the effectiveness of non-interactive teaching is its combination with drawing. Drawing is a visual generative learning activity that requires learners to create external representations of the conceptual and physical structure of the learning material (Ainsworth & Scheiter, 2021; Fiorella & Mayer, 2016). As proposed by Fiorella (2023b), the act of drawing plays a critical role in organizing information meaningfully and may complement the verbal nature of teaching by adding a spatial structure to students' explanations.

In science education, drawing is particularly valuable when learners need to visualize abstract phenomena that are not directly observable—such as optical principles or the particle model (e.g., Cooper et al., 2017). By combining non-interactive teaching with drawing, students may benefit from the distinct but complementary cognitive affordances of both activities: while teaching can foster generalization, drawing can support the re-organization of knowledge (Fiorella, 2023b). Moreover, drawing may facilitate the teaching process itself by providing a visual structure that scaffolds the explanation, potentially leading to higher-quality instructional explanations (Fiorella, 2023b). Together, these processes could promote a more meaningful integration of the learning contents and may help students move beyond surface features and develop deeper, more coherent mental models of the subject matter (Fiorella, 2023b; Rau, 2017).

While the combination of non-interactive teaching and drawing appears promising, it is important to note that previous studies have primarily examined these strategies in isolation (Ainsworth & Scheiter, 2021). Only a few exceptions have investigated the integration of both activities (e.g., Fiorella, 2023a; Fiorella & Kuhlmann, 2020), showing that the combination can potentially boost students' conceptual understanding. However, these studies focused exclusively on laboratory settings with university students and short-term outcomes. As such, it remains unclear whether the potential benefits of combining non-interactive teaching and drawing also apply to inquiry-based authentic classroom settings in schools and whether these benefits can persist over longer periods of time.

1.4.2 Enhancing Non-Interactive Teaching Through Distribution

Another strategy for improving the effectiveness of non-interactive teaching lies in its temporal structure. In many existing implementations, students teach the learning contents only once, typically at the end of a study phase (e.g., Hoogerheide, Visee, et al., 2019; Sibley et al., 2024). However, based on theoretical accounts of generative learning (Fiorella & Mayer, 2016) and findings from interpolated testing (Pan et al., 2024), it is conceivable that distributing non-interactive teaching across multiple points within a study phase may offer cognitive and metacognitive benefits.

From a cognitive perspective, repeated teaching opportunities during a study phase may strengthen the (re-)construction of mental representations by prompting students to revisit, revise, and extend their understanding. Each new generative episode may serve as a chance to reinforce or refine prior processing, supporting knowledge acquisition (Cuddy & Jacoby, 1982; Prinz et al., 2020b). From a metacognitive standpoint, distributing teaching activities across a study phase may provide multiple opportunities for monitoring and self-regulation, allowing students to track their comprehension, adjust strategies, and correct misunderstandings (Lachner et al., 2020; Schleinschok et al., 2017).

Initial research suggests that interpolating a single non-interactive teaching activity during the study phase can be more effective than implementing the activity only after the study phase, likely due to more frequent engagement in monitoring processes (Lachner et al., 2020). However, as shown in the study by Lachner et al. (2020), this finding stems from the only existing study to date that systematically examined the timing of non-interactive teaching—and it did so under controlled laboratory conditions, with university students, with text-based learning materials, and focusing exclusively on a single teaching opportunity. As such, it remains unclear whether multiple distributed teaching episodes offer additional benefits for students' conceptual knowledge and monitoring accuracy—particularly in authentic classroom settings. Moreover, as the study was not situated in an inquiry-based context, its implications for inquiry-based science learning remain uncertain. Furthermore, as no delayed posttest was employed in this study, the question of whether such a distribution also supports lasting learning remains unresolved. These open issues underscore the importance of investigating whether distributing non-interactive teaching across multiple points within a study phase enhances students' conceptual knowledge and monitoring accuracy in inquiry-based authentic school settings.

1.4.3 Enhancing Non-Interactive Teaching With Retrieval Practice

Another promising approach to increasing the effectiveness of non-interactive teaching is its combination with retrieval practice. This learning strategy engages students in actively recalling previously learned information, for instance through quizzes or tests, and has been shown to support the retention of knowledge over time (Karpicke, 2017; Yang et al., 2021). Retrieving information from memory is assumed to reinforce memory traces and contribute to consolidation processes (Karpicke, 2017).

From a theoretical perspective, combining non-interactive teaching and retrieval practice may be especially effective because they serve complementary learning functions: while generative activities such as non-interactive teaching promote knowledge construction, retrieval strengthens consolidation and supports long-term retention (Fiorella, 2023b; Roelle, Endres, et al., 2023). This suggests that combining non-interactive teaching with retrieval practice could support understanding and lasting learning.

However, the two approaches—non-interactive teaching as a generative activity and retrieval practice—have largely been investigated in isolation (see Roelle, Endres, et al., 2023) or in direct comparison (e.g., Jacob et al., 2020). Rather than searching for the single most effective learning task, recent research has shifted toward examining the potential benefits of combining generative learning and retrieval practice, as well as exploring how the sequencing of these tasks may affect learning (Roelle, Endres, et al., 2023). To date, however, there is little empirical research on such sequential combinations, and—to the best of my knowledge—none that explicitly targets non-interactive teaching (for rare examples in related research, see Larsen et al., 2013; Roelle, Froese, et al., 2022). Accordingly, it remains unclear whether non-interactive teaching can be enhanced by sequentially adding retrieval practice, particularly with regard to students' conceptual knowledge and monitoring accuracy in inquiry-based authentic science lessons in school.

1.5 Situating Non-Interactive Teaching Within Contexts of Teaching and Learning

Overall, non-interactive teaching could be a promising approach to support successful subject-specific teaching, as non-interactive teaching—and its potential enhancements through drawing, distributing, and retrieval practice—may foster students' (lasting) learning. It is therefore essential to consider non-interactive teaching within the broader context of teaching and learning in order to develop a comprehensive understanding of how it may contribute to students' learning within classroom teaching.

1.5.1 Non-Interactive Teaching Within Teaching and Learning

Non-interactive teaching, as a generative learning activity, can be aligned with the four key features of teaching proposed by Terhart (2019). First, non-interactive teaching is guided by a clear *pedagogical intention*: students are deliberately prompted to teach learning contents to a fictitious peer with the aim of deepening their understanding and supporting knowledge construction. Second, non-interactive teaching is typically *systematically embedded* within instructional designs, for instance at specific points in an inquiry-based learning phase, indicating its *planned* nature. Third, it occurs *within the formal institutional setting of schooling*, being implemented during formal lessons under the supervision of a teacher. Fourth, although students engage actively in explaining contents, the *professional responsibility* for setting up, guiding, and integrating non-interactive teaching into the overall teaching strategy remains with the *professional teacher*. Consequently, non-interactive teaching can be conceptualized as a structured learning activity within formal teaching.

At the same time, following a cognitively oriented view of learning, the ultimate goal of embedding non-interactive teaching into classroom activities remains to support students' learning: namely, to facilitate the acquisition of new knowledge and the reorganization of existing knowledge structures (Roelle, Lachner, et al., 2023).

1.5.2 Non-Interactive Teaching Within the Offer-Use Model of Teaching

How non-interactive teaching as a generative activity is related to student learning can be considered through the prominent Offer-Use Model of Teaching proposed by Vieluf et al. (2020). In this model, teaching is understood as an offer that students can use as active (co-)constructors of their own learning processes. Therefore, teaching is conceptualized as an interactive process involving teachers, students, and subject matter.

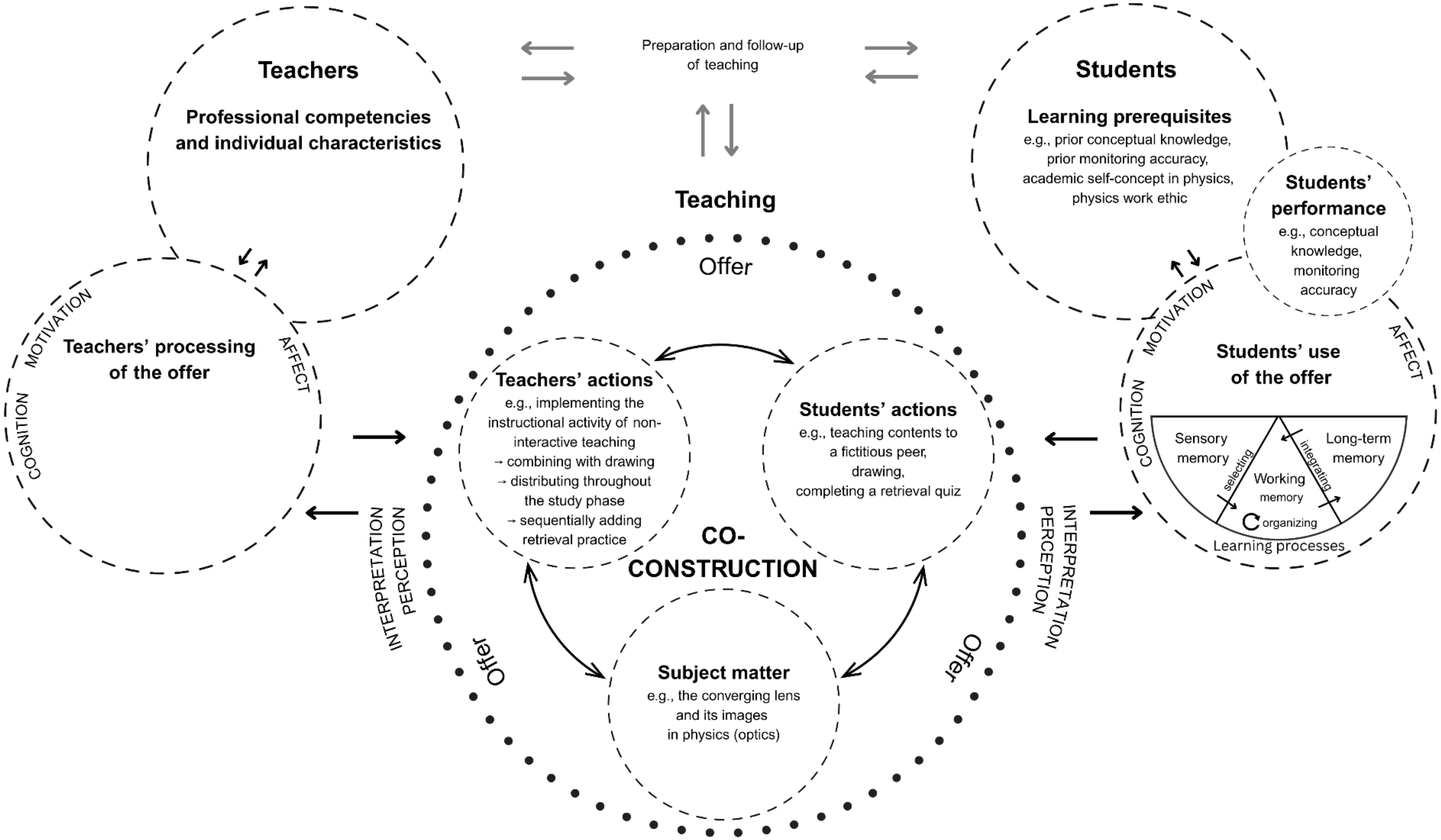
Following the current discourse on teaching (e.g., Keller et al., 2025; Vieluf et al., 2020), it is consistent and established to incorporate a learning perspective informed by teaching and learning research into the Offer-Use Model of Teaching—which originally emerged from teaching quality research. Moreover, the model's authors (Vieluf et al., 2020) have pointed out that its structure is intentionally open and can be specified further depending on the focus of application (e.g., digitalization in school; Syring et al., 2022).

Building on this perspective, the present approach situates non-interactive teaching within the Offer-Use Model of Teaching (Vieluf et al., 2020) as a domain-specific teaching offer. In line with a psychologically oriented understanding of learning, this approach also integrates a model of generative learning (Fiorella & Mayer, 2016). Overall, the focus is placed on teaching and learning processes in relation to non-interactive teaching.

For example, in a physics unit on optics, specifically addressing the topic of "The converging lens and its images", the teacher provides a structured teaching offer that includes a presentation-led introduction to the basic functions of the converging lens and students' experiments, focusing on the effects of partially covering the lens and changing the distance between the object and the converging lens on image formation. To support students in acquiring this new knowledge, the teacher integrates the generative learning activity of non-interactive teaching into the lesson. This learning activity can be implemented in modified formats that combine non-interactive teaching with drawing, distribute it across the study phase, or sequentially add retrieval practice. These design decisions involve both subject-specific and learning-related considerations, including when to implement the activity during the learning process. The students' actions related to non-interactive teaching include teaching the learning contents to a fictitious peer—possibly in combination with drawing—either once or at multiple time points during the study phase, or, after non-interactive teaching, completing a sequential retrieval quiz. In all these cases, non-interactive teaching and its modifications represent a structured opportunity for students to actively engage with the topic of the converging lens and its images. Ideally, students take up the offer, for example, by generating elaborated and complete explanations, which require them to select relevant information from sensory memory, organize it meaningfully in working memory, and integrate it into long-term memory. These internal learning processes reflect the cognitive mechanisms that can underlie successful use of the offer. However, as previous and related research suggests, the quality of students' engagement may depend on additional factors such as task-specific motivation, perceived cognitive effort, or individual learning prerequisites like prior knowledge, academic self-concept, and work ethic. In this sense, non-interactive teaching and its possible modifications can be understood as part of a structured teaching offer that, depending on how it is taken up and used by students, may contribute to students' performance—for example, in terms of conceptual knowledge and monitoring accuracy.

Figure 3

Offer-Use Model of Teaching for Students' Non-Interactive Teaching



Note. Adapted from Vieluf et al. (2020) and Fiorella & Mayer (2016).

2

THE DISSERTATION'S RESEARCH FOCUS

2 The Dissertation's Research Focus

Building on the theoretical and empirical foundations outlined in Chapter 1, the following chapter introduces the specific research focus of this dissertation. It begins by outlining the overarching aim and guiding research question of this work. Subsequently, the three empirical studies that form the core of this dissertation are introduced in relation to their specific contributions to answering the overarching research question. To conceptually situate these studies within the broader context of teaching and learning, the Offer-Use Model of Teaching for Students' Non-Interactive Teaching (see Figure 3) is used to illustrate the positioning of the three studies and their respective contributions to potentially enhance non-interactive teaching in authentic school settings.

2.1 Aims and Research Questions

Supporting all students in learning is a central goal of high-quality science education in school. Non-interactive teaching could therefore be a promising way to support successful teaching. In particular, students with disadvantaged backgrounds and those attending schools in socially deprived areas may benefit from such structured instructional approaches (Bremm et al., 2016; Klein, 2017).

At the same time, while non-interactive teaching shows promise as a generative learning activity, prior research has demonstrated both modest and heterogeneous effects of non-interactive teaching (Kobayashi, 2024; Lachner et al., 2021; Ribosa & Duran, 2022), indicating that this learning activity is not necessarily effective and could benefit from modifications. In addition, most previous studies were conducted in controlled laboratory environments with university students (e.g., Fiorella & Mayer, 2013, 2014; Hoogerheide et al., 2014). These contexts differ substantially from everyday school classrooms, which are typically characterized by features such as greater student diversity, more environmental distractions, and complex social dynamics (e.g., Dinsmore & Alexander, 2012). Thus, it remains unclear whether non-interactive teaching is also effective with school students within the complex context of inquiry-based authentic science education in schools. Furthermore, potential lasting effects of non-interactive teaching remain largely unexplored, as most studies examined only immediate or short-delay learning outcomes (e.g., after one week; Fiorella & Mayer, 2013).

Against this background, the present dissertation aims to systematically investigate how non-interactive teaching can be enhanced to foster students' learning and metacomprehension in inquiry-based authentic school settings, with particular attention to both immediate and lasting learning outcomes. To this end, the overarching research question guiding this dissertation is:

How can the generative activity of non-interactive teaching be enhanced regarding students' conceptual understanding and monitoring accuracy in inquiry-based authentic science lessons—both in terms of immediate and lasting learning?

The approaches outlined in Chapter 1 highlight promising ways to enhance the effectiveness of non-interactive teaching. However, existing research leaves several open questions—particularly regarding the long-term effects of such enhancements and their applicability in inquiry-based authentic classroom settings. To address the overarching research question of this dissertation, three empirical classroom studies were conducted, each investigating a theoretically and empirically grounded potential instructional enhancement of non-interactive teaching. Each study focuses on a distinct instructional strategy, thus contributing uniquely to answering the overarching research question:

1. Study 1: Combining non-interactive teaching with drawing

   Based on generative learning theory (Fiorella, 2023b) and prior research on the benefits of combined non-interactive teaching and drawing (Fiorella, 2023a; Fiorella & Kuhlmann, 2020), it is reasonable to expect that this combination can foster meaningful learning. Against this background, the first study of this dissertation investigated whether combining non-interactive teaching with drawing—as a potential enhancement of non-interactive teaching—supports students' conceptual knowledge and monitoring accuracy in inquiry-based authentic science lessons, regarding both immediate and lasting learning. Moreover, prior research has indicated that students' task-specific motivation (Hoogerheide, Visee, et al., 2019; Jacob et al., 2021) and the quality of students' explanations can influence learning (Fiorella & Kuhlmann, 2020; Jacob et al., 2020; Lachner et al., 2018). Therefore, the first study also examined whether these variables contributed to the observed learning effects.

2. Study 2: Distributing non-interactive teaching throughout a study phase

   Drawing on generative learning theory (Fiorella, 2023b) and interpolated practice (Pan et al., 2024), the second study investigated whether distributing non-interactive teaching across multiple points of an inquiry-based study phase—as a potential enhancement of non-interactive teaching—supports students' conceptual knowledge and monitoring accuracy in authentic science lessons, regarding both immediate and lasting learning. Following Lachner et al. (2022), the study additionally accounted for students' individual differences that may shape learning. In line with prior research, prior knowledge (Hoogerheide, Renkl, et al., 2019) and academic self-concept (Jacob et al., 2022) were explored. As an extension, work ethic as a facet of conscientiousness was considered, given its documented relevance for learning in related research (Bareis et al., 2024; Song et al., 2020; Spielmann et al., 2022).

3. Study 3: Combining non-interactive teaching with retrieval practice

   Generative learning such as non-interactive teaching supports the construction of meaningful mental representations (Fiorella, 2023b), while retrieval practice supports consolidation and long-term retention (Karpicke, 2017). Due to their distinct yet complementary functions, combining generative activities with retrieval practice may enhance learning outcomes (Roelle, Endres, et al., 2023). Accordingly, the third study investigated whether combining non-interactive teaching and retrieval practice—as a potential enhancement of non-interactive teaching—improves students' conceptual knowledge and monitoring accuracy in inquiry-based authentic school settings, regarding both immediate and lasting learning. Given that students' engagement in non-interactive teaching may shape the effectiveness of subsequent retrieval practice (e.g., Roelle, Froese, et al., 2022), we additionally explored associations between generation, retrieval, and explanation quality (i.e., completeness, elaboration, correctness).

2.2 Locating the Research Focus Within the Offer-Use Model of Teaching for Students' Non-Interactive Teaching

To situate the research focus of the dissertation within the broader context of teaching and learning, this section uses the Offer–Use Model of Teaching for Students' Non-Interactive Teaching (Figure 3) to provide a conceptual framing of the three classroom studies and their respective contributions (see Figure 4).

1. Study 1 centers on the teaching offer, investigating whether combining non-interactive teaching with drawing may improve students' conceptual knowledge and monitoring accuracy. It also considers learner-level factors such as task-specific motivation and explanation quality, thereby linking the offer to its actual use.

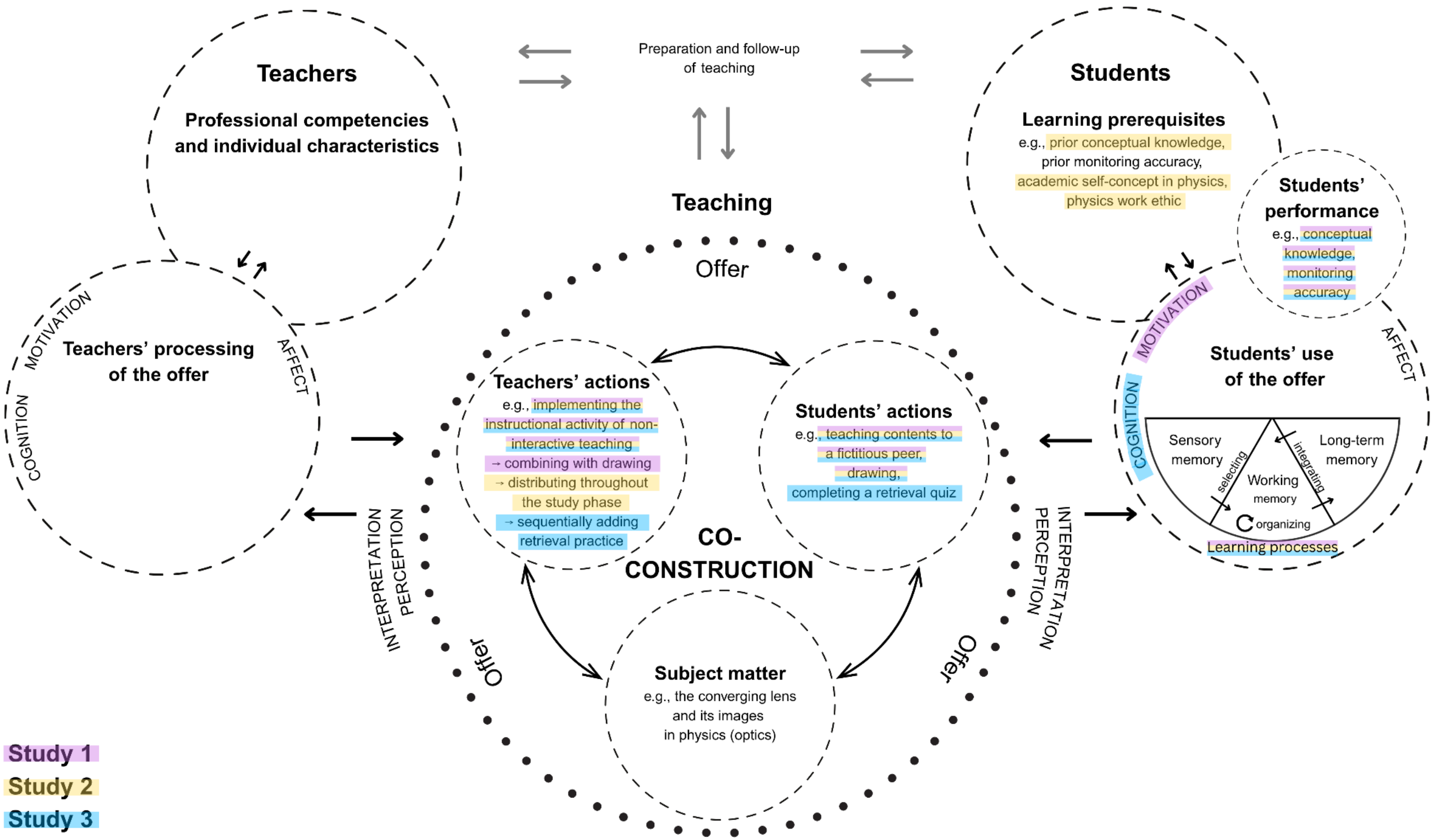
2. Study 2 focuses on the timing of the offer by distributing non-interactive teaching across the study phase. Additionally, it addresses student prerequisites (i.e., prior knowledge, academic self-concept, work ethic) that may shape students' performance (i.e., conceptual knowledge, monitoring accuracy).

3. Study 3 investigates how non-interactive teaching can be enhanced by sequentially adding retrieval practice. It thereby focuses on the offer itself, investigating its influence on students' conceptual knowledge and monitoring accuracy. In doing so, it explores the relationship between the offer and students' use of it with respect to their learning processes.

Taken together, the three classroom studies could provide differentiated insights into how non-interactive teaching—as a generative learning activity—can be modified and potentially enhanced regarding students' conceptual knowledge and monitoring accuracy in authentic instructional contexts, thereby directly addressing the overarching research question of this dissertation.

Figure 4

The Dissertation's Research Focus located in the Offer-Use Model of Teaching for Students' Non-Interactive Teaching



Note. Adapted from Vieluf et al. (2020) and Fiorella & Mayer (2016).

3

STUDY 1

COMBINING NON-INTERACTIVE TEACHING AND DRAWING FOSTERS CONCEPTUAL KNOWLEDGE BUT NOT MONITORING ACCURACY FROM GUIDED INQUIRY IN SCIENCE LEARNING

Russ, H., Sibley, L., Flegr, S., Kuhn, J., Hoogerheide, V., Scheiter, K., & Lachner, A. (in press). Combining non-interactive teaching and drawing fosters conceptual knowledge but not monitoring accuracy from guided inquiry in science learning. *Journal of Educational Psychology.* https://doi.org/10.1037/edu0000971

3 Study 1: Combining Non-Interactive Teaching and Drawing Fosters Conceptual Knowledge but not Monitoring Accuracy From Guided Inquiry in Science Learning

Abstract

Although inquiry-based learning is a widespread instructional strategy in science education, students often tend to process the contents superficially, which can hamper learning. In this classroom experiment, we investigated whether the generative activity of non-interactive teaching, in which students verbally explain contents to a fictitious, non-present peer, can be enhanced by drawing to support students' meaningful learning from guided inquiry. The study was conducted in an authentic physics education context in schools with secondary students (N = 590), contrasting three generative conditions (teaching-only, teaching + provided visualization, teaching + drawing) to a restudy control condition. Their potential effects were examined on immediate and lasting conceptual knowledge and monitoring accuracy, defined as students' metacognitive ability to judge their own performance. Regarding performance on the immediate conceptual knowledge test, generative activities were more effective than restudy. Also, in line with our expectations, the combined visualization conditions (teaching + visualization, teaching + drawing) outperformed teaching-only; moreover, teaching + drawing was more effective than teaching + visualization. The latter finding could be traced back to the fact that creating an additional drawing increased students' task interest and the completeness of their verbal explanations. Regarding monitoring accuracy, no effects were found. Contrary to expectations, there were no significant differences among conditions on lasting (meta-)cognitive learning as assessed eight weeks after the lesson. The results demonstrate that combining generative activities enhances meaningful inquiry learning and highlights the need to further enhance generative activities to aid lasting learning.

Educational Impact and Implications Statement

This classroom study demonstrates that teaching a fictitious peer and generating drawings can help secondary students better understand science concepts. By embedding these strategies into curriculum-aligned, inquiry-based authentic lessons, the study underscores their potential to make classroom learning more engaging and effective. These findings provide valuable insights for educators aiming to enhance learning outcomes through practical and scalable teaching methods.

Keywords

Generative learning, inquiry learning, non-interactive teaching, drawing, lasting learning

3.1 Introduction

Scientific literacy is essential for academic success and students' participation in today's knowledge society (OECD, 2016). However, recent large-scale achievement studies (e.g., Mullis et al., 2020; OECD, 2023) have highlighted that students often demonstrate low levels of scientific literacy. Inquiry-based activities are regarded as fostering scientific literacy, since they require students to mimic the professional activities of researchers under supervised conditions (Pedaste et al., 2015). Despite the potential of inquiry-based activities, empirical research indicated that learning from inquiry often puts high (meta-)cognitive demands on students, which can preclude the envisioned benefits regarding lasting knowledge acquisition (Alfieri et al., 2011). Thus, students' learning often remains on a superficial level without students grasping the underlying concepts and principles (Goldwater & Schalk, 2016; Schneider et al., 2011).

Based on generative learning theory (Fiorella, 2023b), it can be assumed that generative activities can help students actively make sense of the new information in inquiry-based settings, resulting in higher (meta-)cognitive learning outcomes (Chi & Wylie, 2014). For instance, verbal activities, such as teaching the learning contents to a fictitious peer (cf. learning by non-interactive teaching, Lachner et al., 2022), may help elaborate and generalize the learning contents. Generating drawings, however, may be particularly effective to re-organize the contents (Ainsworth & Scheiter, 2021; Fiorella, 2023b; Fiorella & Mayer, 2016; Lachner et al., 2021; Rau, 2017).

Whether combining verbal and visual activities aids learning from inquiry is yet an open question, as previous research has so far treated these activities in isolation (Ainsworth & Scheiter, 2021; Lachner et al., 2021) or only included self-contained expository learning materials in rather small-scale laboratory settings with university students as participants (Fiorella, 2023a; Fiorella & Kuhlmann, 2020). Additionally, prior research predominantly focused on immediate rather than lasting learning effects. Even when delayed tests were conducted, they were often limited to just one week later. However, lasting learning is crucial in education (OECD, 2016; T. Richter et al., 2022), making it essential to examine whether effects persist over longer periods of time.

Against this background, we conducted a large-scale classroom experiment in inquiry-based authentic physics lessons with secondary school students to investigate whether combining teaching and drawing can enhance students' (meta-)cognitive learning from inquiry. Moreover, we investigated whether potential improvements are due to the presence of a visualization or due to the act of generating a drawing oneself. In order to investigate whether combining non-interactive teaching and drawing results in lasting learning effects, we also administered a delayed posttest after eight weeks.

3.1.1 Enhancing Conceptual Understanding in Science Education Through Inquiry Learning

Supporting conceptual understanding that goes beyond the observable phenomena is a central focus in science education (Chi & Roscoe, 2002; de Jong et al., 2023; Schneider et al., 2011). Inquiry-based learning, in which students mimic professional research activities under supervised conditions, is regarded as helping students develop conceptual understanding of the phenomena under investigation (de Jong, 2019). Inquiry-based learning commonly comprises four inquiry phases, forming a self-contained inquiry cycle: orientation, conceptualization, investigation, and conclusion (Pedaste et al., 2015). For instance, first, students are introduced to a topic such as "the converging lens and its images" by giving them the opportunity to make a real-world observation of the phenomenon and are provided with prerequisite knowledge about the topic (orientation). Second, students are asked to generate hypotheses, in our example relating to the imaging process caused by a converging lens (conceptualization). Third, they conduct experimental trials to test their hypotheses (investigation). When conducting the experiments, the students collect data and document their experiments, for example, by filling in a table with their observed results. Fourth, the students form a data-based conclusion, compare their hypothesis with their final findings and discuss their results (conclusion, Pedaste et al., 2012, 2015).

Despite the potential benefits of inquiry for conceptual learning, empirical research has indicated that inquiry often puts high (meta-)cognitive demands on students, which can constrain conceptual learning (Alfieri et al., 2011; Kant et al., 2017; Kirschner et al., 2006). Thus, inquiry learning often remains on a superficial level without sufficiently helping students to understand underlying concepts or principles (Goldwater & Schalk, 2016; Schneider et al., 2011). For instance, Kant et al. (2017) examined inquiry-based tasks versus learning by video-modeling examples with 107 seventh-grade students. Results showed that students who engaged in inquiry-based tasks answered fewer prompts during instruction correctly than students who watched a video-modeling example ($\eta_p^2 = .18$). Moreover, regarding monitoring accuracy, the authors showed that solving two inquiry-based tasks led to underconfidence, while watching two video-modeling examples resulted in accurate judgments. Additionally, students who first solved an inquiry-based task before watching a video-modeling example showed lower learning outcomes than when learning in the reversed order ($\eta_p^2 = .14$).

Meta-analyses have also consistently shown the challenging nature of inquiry-based learning, particularly for novices and for less guided inquiry-based tasks (Alfieri et al., 2011; Lazonder & Harmsen, 2016). Rather than demonizing inquiry-based learning, these findings suggest that students require additional assistance to make sense of the processes of inquiry-based learning for meaningful conceptual understanding (Zacharia et al., 2015).

3.2 Generative Activities to Foster Inquiry-Based Learning

In line with Wittrock's generative model of learning (1989, 2010) and Fiorella's generative sense-making framework (2023b), incorporating generative activities in inquiry-based learning environments can enhance students' ability to actively understand new learning contents. This process includes helping students make sense of their experimental findings and the relationships between the variables analyzed during the experiments (Zacharia et al., 2015). By selecting relevant information from the learning contents, organizing it in working memory, and integrating the newly generated knowledge with prior knowledge, students can construct meaningful mental representations. These representations may also be applied to different problem-solving contexts (Fiorella & Mayer, 2016; Wittrock, 1989).

During their active engagement in generative activities, students may generate ideas that go beyond the given or observed information, which may result in elaboration and a constructive mode of learning (Chi & Wylie, 2014; Van Meter & Firetto, 2013). In addition to supporting cognitive processes, generative activities may also contribute to students' metacognitive processes, such as monitoring their current understanding, as externalizing one's knowledge allows students to reflect on what they know and what they do not know. Detecting potential gaps in their understanding could enable students to resolve potential comprehension problems, provided they are given the opportunity to address and repair these gaps, thereby fostering more effective regulation of their learning process (Fiorella & Mayer, 2016; Roscoe, 2014). Assisting accurate monitoring is particularly relevant because students tend to often overestimate their current understanding of a given topic, which could impair students' learning (Dunlosky & Rawson, 2012; Prinz et al., 2020a; Roscoe, 2014). By prompting students to engage in generative activities, such as teaching concepts to a fictitious peer or creating visual representations, these strategies may encourage them to critically evaluate their knowledge and become more aware of their learning (Ainsworth & Scheiter, 2021; Fiorella, 2023b; Fiorella & Mayer, 2016). This increased awareness can help mitigate overconfidence and improve the process and outcomes of learning.

These cognitive and metacognitive generative processes could contribute to deeper cognitive engagement during inquiry—apart from physically interacting with experimental materials (de Jong, 2019; Mayer, 2004), possibly resulting in a rich and lasting conceptual understanding (Barnett & Ceci, 2002; Krajcik & Delen, 2017; Roscoe, 2014). Following Fiorella (2023b), teaching activities (e.g., non-interactive teaching) and visualizing activities (e.g., drawing) are two prominent generative activities. Each activity is regarded as having a unique and complementary cognitive affordance: Whereas asking students to teach may trigger them to generalize and elaborate the previously learned contents, visualizing may help to re-organize and structure such contents (Fiorella, 2023b; Rau, 2017; Van Meter & Firetto, 2013). Thereby, it can be argued that visualizing may be a facilitator to enhance the quality of teaching (Cromley, 2020; Fiorella, 2023b; Fiorella & Jaeger, 2023; Fiorella & Kuhlmann, 2020).

3.2.1 Learning by Non-Interactive Teaching

Learning by non-interactive teaching is considered an effective verbal generative activity for enhancing students' learning (e.g., Fiorella & Mayer, 2014, 2016; Hoogerheide et al., 2014). During non-interactive teaching, students provide a verbal instructional explanation of the previously learned contents to a fictitious peer (Lachner et al., 2022). Non-interactive teaching may trigger distinct cognitive and metacognitive processes, which in turn should yield better (meta-)cognitive learning outcomes (Fiorella, 2023b; Fiorella & Mayer, 2016; Jacob et al., 2020). The benefits of teaching may depend on the quality of the generated explanations, such as their completeness in addressing concepts (e.g., Jacob et al., 2020) or the number of elaborations provided (e.g., Fiorella & Kuhlmann, 2020). Additionally, from a psychological perspective (Lachner et al., 2022), non-interactive teaching requires students to tailor an instructional teaching to a fictitious peer which may increase students' motivation (e.g., interest), a factor shown to enhance learning (Hoogerheide, Visee, et al., 2019; Jacob et al., 2021).

Fiorella and Mayer (2013) investigated the effectiveness of non-interactive teaching through two experiments. University students studied materials either without the intention to teach the learning contents before taking a comprehension test (control group) or with the intention to teach. Among those expected to teach, some immediately took the comprehension test (preparation group), while others taught the learning contents to fictitious others (teaching group) before taking the test. Both teaching expectancy groups showed better immediate learning outcomes than the control group (teaching vs. control: $d = 0.82$; preparation vs. control: $d = 0.59$; experiment 1). However, only the actual teaching group showed better retention in a one-week-delayed posttest (teaching vs. control: $d = 0.79$; preparation vs. control: $d = 0.24$; experiment 2), indicating that the actual act of teaching was essential for students' lasting learning.

In addition, a few studies showed that non-interactive teaching may also contribute to students' monitoring accuracy. For example, in an experiment by Fukaya (2013), university students studied the learning materials with the intention to teach and then either taught the contents (teaching condition) or accomplished a knowledge test (expect-to-teach condition). Students in the control condition read the texts with the intention to provide keywords. Before answering the posttest, students self-assessed their actual knowledge by rating how many points they would achieve in the posttest (judgment of learning). Results showed a main effect of non-interactive teaching on students' monitoring accuracy ($\eta^2 = .17$; see Jacob et al., 2020; Lachner et al., 2020, 2021, for non-significant findings). Relatedly, several studies demonstrated that non-interactive teaching also has positive effects on student motivation (Hoogerheide, Visee, et al., 2019; Jacob et al., 2021). Apparently, non-interactive teaching increases task-specific motivational orientations, which contributes to increases in learning outcome.

The benefits of non-interactive teaching were also reflected in a recent meta-analysis by Kobayashi (2024), who demonstrated a small yet significant effect of non-interactive teaching ($g = 0.27$), based on 39 experimental studies (see also Lachner et al., 2021, for similar evidence). A considerable number of these prior studies have incorporated non-interactive teaching as a retrieval activity (e.g., Fiorella & Mayer, 2013; Jacob et al., 2021), meaning that students did not have access to the study materials during non-interactive teaching. This *closed-book approach* (see Roelle, Endres, et al., 2023) inherently involves a retrieval effect, making it difficult to isolate the generative aspects of learning. However, this approach ensures that students rely on their own knowledge and understanding of the topic, rather than simply reading or copying from available materials. By contrast, an *open-book approach* allows students to consult materials during teaching, providing an opportunity to address knowledge gaps (e.g., Sibley et al., 2022; see also Hiller et al., 2020; Roelle, Endres, et al., 2023). While this method does not explicitly encourage retrieval, it supports self-regulated learning by enabling students to refine their understanding as they engage in teaching. However, it also carries the risk that students may rely excessively on the provided materials, potentially diminishing the generative learning processes. To date, one study has demonstrated that learning with an open-book approach can be more effective (Sibley et al., 2022). Nevertheless, the conditions under which open- or closed-book approaches are most beneficial remain unclear, warranting further investigation.

Moreover, most of the studies on non-interactive teaching were laboratory studies with self-contained study material, making it difficult to gauge whether the finding may generalize to authentic classroom settings with high levels of problem-orientation, such as learning from inquiry. One exception is the study by Jacob et al. (2022). In this study, seventh graders first received direct instruction on the concepts of photosynthesis and then conducted a virtual experiment in which they investigated the effects of light intensity, temperature, and carbon dioxide level on oxygen flow. Afterwards, the students either taught a fictitious peer by generating a written or oral explanation, or they retrieved the information through an unguided open recall task. The authors found that particularly students with low levels of academic self-concept profited from non-interactive teaching (b = 0.49). Overall, the findings demonstrate that non-interactive teaching has potential but might not be that effective on its own concerning learning from inquiry.

3.2.2 Learning by Drawing

Learning by drawing is a visualizing generative learning activity in which students generate external visual representations to communicate the learning material's physical and conceptual structure (Ainsworth & Scheiter, 2021; Fiorella, 2023b; Fiorella & Mayer, 2016; Wu & Rau, 2019). The key function of drawing lies in the meaningful organization of the learning contents (Fiorella, 2023b). Typically, the benefits of drawing are tied to the accuracy of drawings produced by students as an indicator of how well they elaborated on the contents (cf. prognostic drawing effect, Schwamborn et al., 2010; see also Scheiter et al., 2017). Although drawings can be used to simply re-represent what is being observed (e.g., sketching the set-up of an experiment in physics), many uses of drawing are constructive, whereby the visuospatial representation goes beyond the given information (Ainsworth & Scheiter, 2021).

For example, in science education, Cooper et al. (2017) showed how model-based reasoning was improved when students were asked to draw what they observed in a chemistry experiment (e.g., a fluid changing its color after adding a solvent) at a more abstract, conceptual level (e.g., referring to the particle model of matters). In the review by Fiorella and Zhang (2018), constructive drawing (compared to reading text or text-focused strategies) yielded medium (d = 0.40; comprehension) to large (d = 0.70; transfer) effect sizes (see also Cromley et al., 2020).

Drawing has not only been shown to be effective in improving students' cognitive but also metacognitive processes, as it helps students recognize and possibly close knowledge gaps (Van Meter & Firetto, 2013). In a recent study, Fiorella and Jaeger (2023) explored how different instructional visual formats affect students' monitoring accuracy. University students studied biology texts either using the text alone, with provided visualizations, with generating their own visualizations, or with animations of generated visualizations by an instructor. The authors found that students who generated visualizations or received animations of instructor-generated visualizations showed better monitoring accuracy than students who were either provided with visualizations or received no visualizations (d = 0.36, for teaching judgments).

To date, little is known about whether drawing enhances learning due to its constructive nature, and the available evidence is mixed. Some findings indicate that self-generated drawings are more effective than provided visualizations (Mason et al., 2013), while others show no differences (Hall et al., 1997; Schmidgall et al., 2019; Schwamborn et al., 2011), or even indicate that provided visualizations lead to better learning outcomes (Fiorella, 2023a). Whether students benefit more from provided visualizations or generating their own drawings during learning remains an open question.

For a comprehensive view on drawing, it is also worth noting that previous research has shown that drawing is not always an effective or appropriate learning strategy (Ainsworth & Scheiter, 2021; Brod, 2021; Fiorella & Zhang, 2018). Drawing can impose high cognitive demands on students, requiring them to select, organize, and transform information, such as from linear texts into visuospatial representations. Additionally, they must compare and refine their drawings to align with the original information—a process that also depends on metacognitive monitoring and regulation (Ainsworth & Scheiter, 2021; Fiorella, 2023b; Mayer et al., 1995; Schwamborn et al., 2010; Van Meter & Garner, 2005). This may be particularly challenging for younger students, whose executive functions, such as working memory, attention, and cognitive control, are still developing (Brod, 2021; Diamond, 2013).

Evidence for younger students is limited but shows that they often fail to benefit from drawing, even with instructional support (e.g., Rasco et al., 1975; Van Essen & Hamaker, 1990; Van Meter et al., 2006) in contrast to older students (e.g., Fiorella & Kuhlmann, 2020; Schmeck et al., 2014; for a comprehensive review, see Brod, 2021). Additionally, students with low self-efficacy for drawing may struggle to participate actively in the task, reducing its potential benefits (Y. Wang & Zhang, 2024). Moreover, the time-intensive nature of drawing might limit its feasibility in time-constrained classroom environments (Schmidgall et al., 2019).

3.2.3 Combining Non-Interactive Teaching and Drawing

To increase generative processing in inquiry-based learning settings, combining non-interactive teaching with drawing could enhance learning, which may hold true especially for visuo-spatial learning tasks such as experimenting in science. Combining these generative activities could support the meaningful integration of new information by promoting both generalization and re-organization processes, which should help learners move beyond the phenomenological level to develop a deeper conceptual understanding (Fiorella, 2023b). According to Fiorella's generative sense-making framework (2023b), drawing can serve to facilitate teaching and thereby may help students to generate higher quality explanations.

Interestingly, most prior studies have addressed non-interactive teaching and drawing in an isolated manner (Ainsworth & Scheiter, 2021). To date, only Fiorella and Kuhlmann (2020) and Fiorella (2023a) examined the combination of teaching and drawing. In the experiment of Fiorella and Kuhlmann (2020), university students (N = 120) either taught the contents (i.e., human respiratory system) to a fictitious peer, created drawings, drew and taught, or restudied the materials. The authors found a main effect of teaching in a one-week-delayed posttest. Moreover, students who drew and taught outperformed the drawing-only (d = 0.65), the teaching-only (d = 0.99), and the restudy condition (d = 1.46). In a follow-up experiment by Fiorella (2023a) following a 2 × 2 design, university students (N = 132) taught with generating words, taught and drew, taught with provided visualizations, or taught with provided words. Results indicated that teaching with provided visualizations or drawing was better than teaching with provided words or generating words in a transfer test (η^2 = .04). Moreover, teaching with provided visualizations resulted in higher drawing test performance than teaching and drawing (η^2 = .04).

Overall, the aforementioned results suggest that the combined use of learning by non-interactive teaching and drawing can potentially boost students' conceptual understanding. However, previous studies only investigated effects after a one-week-delay, leaving open whether such benefits persist over a longer period, which is crucial for examining how lasting learning can be achieved. Additionally, Fiorella and Kuhlmann (2020) did not focus on the different affordances of drawing, such as visualization and the act of drawing itself, and Fiorella (2023a) found only partial support that provided visualizations and drawing facilitate teaching. Moreover, whether the findings generalize to inquiry-based authentic settings in school is still an open question. These unresolved issues highlight the need for further research to more thoroughly investigate the immediate and lasting effects of combined non-interactive teaching and drawing and their applicability in real-world inquiry-based school environments, making this a key focus for the present study.

3.3 The Present Study

Based on the previous considerations, we conducted a classroom experiment with seventh and eighth-grade secondary school students, attending an inquiry-based authentic physics unit on the topic of converging lenses (geometrical optics). We investigated whether combining students' non-interactive teaching to a fictitious peer and drawing during inquiry-based learning enhances their cognitive (i.e., conceptual knowledge) and metacognitive (i.e., monitoring accuracy) learning, regarding both immediate and lasting learning after eight weeks. In addition, we tested whether such a potential combinatory effect is a result of the visualization of the learning contents or of the active generation of drawings (cf. Fiorella, 2023a; Schmidgall et al., 2019). Further, we explored whether task-specific motivational mechanisms underly a potential effect of combining students' non-interactive teaching and drawing and additionally analyzed students' explanation characteristics. We tested the following preregistered hypotheses (https://aspredicted.org/R35_NP8):

3.3.1 Generation Hypothesis

Following the generative learning theory (Fiorella, 2023b; Fiorella & Mayer, 2015; Wittrock, 1989, 2010), we hypothesized that students in the generative conditions (teaching-only, teaching + visualization, teaching + drawing) outperform students in the restudy control group regarding their immediate and lasting a) conceptual knowledge (Fiorella & Mayer, 2013, 2014; Hoogerheide, Visee, et al., 2019) and b) monitoring accuracy (Fukaya, 2013; Jacob et al., 2020).

3.3.2 Visualization Hypothesis

Regarding the generative sense-making framework (Fiorella, 2023b), we hypothesized that students who teach the learning contents with an added visualization component (teaching + visualization, teaching + drawing) outperform students in the teaching-only condition regarding their immediate and lasting a) conceptual knowledge (Cooper et al., 2017; Schmidgall et al., 2019) and b) monitoring accuracy (Fiorella & Jaeger, 2023).

3.3.3 Drawing Hypothesis

Previous research demonstrated that student-generated drawings (Cooper et al., 2017; Fiorella & Kuhlmann, 2020) or provided visualizations (Ainsworth & Loizou, 2003; Cromley et al., 2010) can support students' meaningful learning. Based on these findings and the generative learning theory (Fiorella, 2023b), we hypothesized that students who teach and draw (teaching + drawing) outperform students who teach while having access to a provided visualization (teaching + visualization) regarding their immediate and lasting a) conceptual knowledge (Ainsworth & Loizou, 2003; Fiorella & Kuhlmann, 2020) and b) monitoring accuracy (Fiorella & Jaeger, 2023; Schleinschok et al., 2017).

Following Fiorella's generative sense-making framework (2023b) and the findings of prior research (e.g., Fiorella, 2023a; Fiorella & Jaeger, 2023; Fiorella & Kuhlmann, 2020), we assumed that the (meta-)cognitive effects of our interventions are more pronounced in the eight-week delayed conceptual knowledge test than in the immediate test.

3.3.4 Further Explorative Analyses

Prior research indicated that teaching increased students' task-specific motivation (Hoogerheide, Visee, et al., 2019; Jacob et al., 2021), and that the characteristics of students' explanations determine students' learning (Fiorella & Kuhlmann, 2020; Jacob et al., 2020; Lachner et al., 2018). We therefore explored whether students' task-specific motivation and the characteristics of students' explanations contributed to potential (meta-)cognitive effects.

3.4 Method

3.4.1 Participants and Design

In total, thirty seventh and eighth grade classes from eleven secondary schools in south-west Germany participated in the study ($N = 720$). We excluded data from participants who did not attend the main part of the study ($n = 130$), which took place on the second of three scheduled dates at the schools and included a topic-related teaching unit, the interventions, and the immediate posttest. This resulted in a final sample size of *N* = 590 school students. Our actual sample size surpassed the required minimum of 206 students by large, which was determined through an a-priori power analysis with G*Power ($f = 0.25$, $\alpha = .05$, $1\text{-}\beta = .90$; ANCOVA). The larger sample size resulted from unexpectedly high interest in our study, as 11 out of 13 schools agreed to participate due to recognized benefits for their students. Additionally, we received an unusually high return rate of consent forms from the students' legal guardians (over 90%). The increased sample size offers significant advantages by enhancing the reliability of our findings and enabling the detection of smaller effects, thus making a more substantial contribution to the field. The mean age of the participants was 12.79 years (*SD* = 0.70), and 49.91% were female. The majority of the students had German as their first language (65.69%), 23.30% of the participants had a native language other than German, and 11.01% were bilingual with German.

The classroom experiment was implemented in an authentic physics teaching context in school, included a pre-post- and eight-week-delayed test, and had a one-factorial between-subjects design with four conditions. Students were randomly assigned to one of the four experimental conditions within each class. Students either taught the learning contents to a fictitious peer (teaching-only, *n* = 144), taught with a provided visualization (teaching + visualization, *n* = 142), taught in combination with drawing (teaching + drawing, *n* = 154), or restudied the learning contents (restudy, *n* = 150). Restudying was realized as a control condition because this learning activity is generally not regarded as promoting generative processing (Fiorella & Mayer, 2016). Students of all conditions had access to the learning materials during the learning activities (teaching-only, teaching + visualization, teaching + drawing, restudy). This open-book format was realized to avoid a "hidden" retrieval effect (Roelle, Endres, et al., 2023; Sibley et al., 2022).

3.4.2 The Teaching Unit

The teaching unit was about the converging lens and its images (physics, geometrical optics) and was curriculum-aligned within physics education. It was taught across all participating classes by the first author, a certified physics teacher with 10 years of teaching experience, who also personally conducted all data collection. All learning materials (introduction, paper-based overview sheet, paper-based experimentation worksheet) were based on a previous study by Flegr et al. (2023) and were carefully adapted to the purpose of this study by the first author.

3.4.2.1 Introduction to the Converging Lens

The introduction to the converging lens and its basic functions included lenses in everyday life, possible linking the students to their previous knowledge by prompting them with questions such as "Where can you find lenses in your everyday life?", followed by a class discussion and joint examination of example images. Moreover, the introduction offered a definition of the converging lens and central terminology, an explanation and illustration of the path of light that refracts through a converging lens, and basic information on resulting formed images. Students received an overview sheet of the taught basic knowledge about the converging lens and its functions (see first worksheet at Flegr et al., 2023).

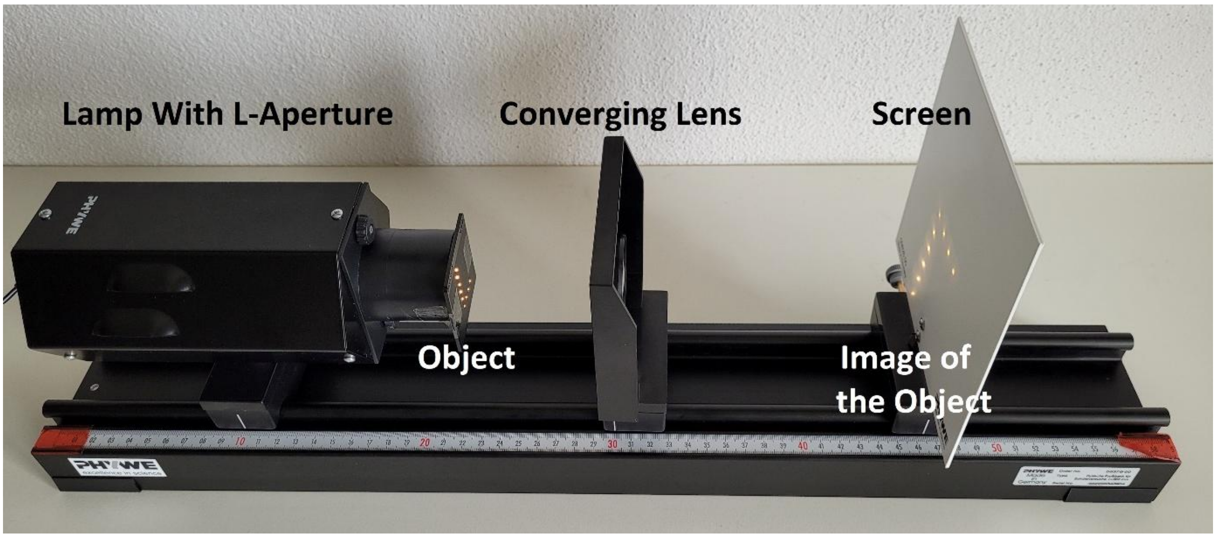
3.4.2.2 Students' Experiments

Students subsequently explored the effects of a partially covered converging lens (diameter of the lens) and the distance between object and converging lens on the resultant image of the object. For these experiments (Figure 5) students used an optical bench, an LED lamp with a "Perl-L"-aperture, a converging lens, and a white screen. The light source creates a bright 'L'-shaped object, which can be positioned closer or farther away from the converging lens. The light from the object ('L') is refracted through the converging lens and forms an image on a moveable screen on the other side of the lens ('L' is now swapped left-right and top-bottom).

An experimentation worksheet guided the students during the experiments, including hypothesis formulations (e.g., "What do you suppose happens to the image of the object when the lens is partially covered?"), multiple-choice questions (e.g., "Is the complete 'L' still imaged when the lens is half covered? Yes, there is no difference. / Yes, but the 'L' on the screen is not as bright as before. / No, the 'L' is cut off. / No, the 'L' is no longer visible"), an open-question ("Can you explain why that is?"), a table for documenting the observation results (including object distance, image distance, and characteristics of the image), mnemonic sentences from which the correct word elements were to be selected (e.g., "If the lens is partly covered, an image of the complete object *is still formed / is not formed* on the screen"), and a hypotheses comparison with the final findings by the students (e.g., "Compare the mnemonic with your previously formulated hypothesis 1. Was your assumption different than the result? Yes, completely different. / A bit. / No, it was the same").

Figure 5

Experimental Setup on the Converging Lens and its Images



3.4.3 Learning Activity

During the learning activity, students either taught the learning contents to a fictitious peer (teaching-only), taught with a provided visualization (teaching + visualization), taught and drew a visualization (teaching + drawing), or restudied the contents. Students in the generative conditions (teaching-only, teaching + visualization, teaching + drawing) received the following written instruction:

In a message, the student ***Mia*** *wrote to you* ***two assumptions*** *about the converging lens and its images.* ***Reply to Mia*** *by creating a* ***clear and detailed voice message to Mia*** *so that she can understand the contents without any additional information. Respond to* ***Mia's assumptions****. You are allowed to use the materials of the topic, but it is very important that you* ***DO NOT*** *read from them when recording the voice message but* ***formulate it in your own words*** *and incorporate your* ***own thoughts***[3]*. You have a total of* ***15 minutes*** *to complete this task. Be sure to use all the time.*

In the teaching-only condition, students were presented with a mock-up chat with the fictitious peer Mia (Figure 6, left side). In this chat, Mia expressed two assumptions: first, "I actually believe: When the lens is covered half, only half of the L is shown as an image.", and second, "I actually believe: When I move the object towards the lens, I also have to move the screen closer to the lens in order to see a sharp image on the screen." These assumptions address two core misconceptions that students have when learning about image formation by a converging lens (see Wörner et al., 2022). Students had to teach the learning contents to Mia by recording a voice message.

The mock-up messenger was also presented to the students in the teaching + visualization condition including Mia's two assumptions supplemented with two physically

[3] During the coding process, we checked students' adherence to these instructions and found no evidence that students read from or copied the materials while teaching or drawing.

correct visualizations of the learning contents and Mia's request to teach the topic to her with the help of the visualizations (Figure 6, middle).

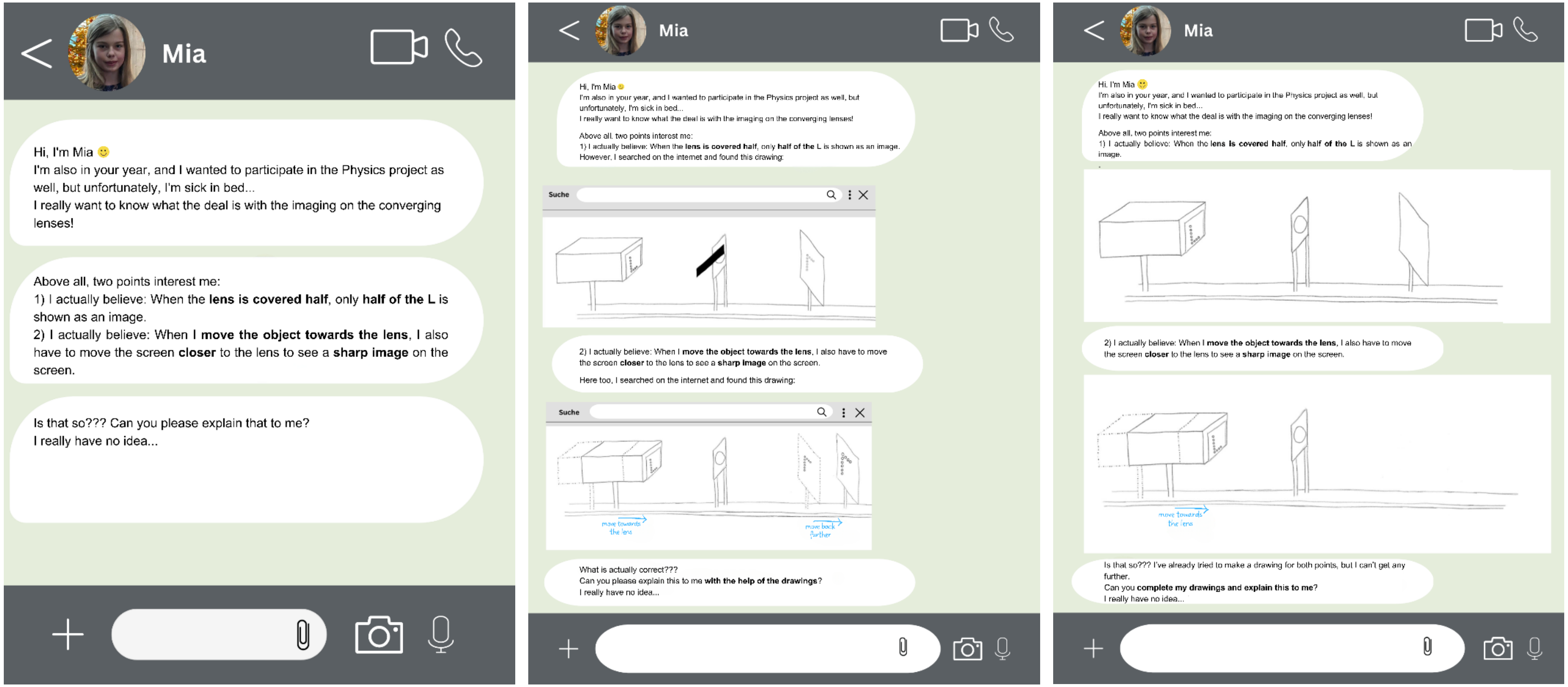
Students of the teaching + drawing condition also received Mia's message and had to teach the learning contents to her by recording a voice message. Moreover, students were instructed to draw using two template pictures during teaching as a means to explain the contents to Mia (Figure 6, right side).

To ensure that students in the control group focused on the same contents, we supplied them with detailed instructions that also addressed both misconceptions:

Restudy *the contents of today's physics lessons. To do this, use* ***today's physics materials****: You have received an overview sheet on the basics about the converging lens and its images and you have the instructions and results of the experiments. Use these materials for restudying.* ***Focus on the following two points:*** *What happens to the image if the lens is partially covered? What happens to the image if the distance between the object and the lens is changed? Take* ***notes*** *on this sheet. You may use the front and back of the sheet for this. You have a total of* ***15 minutes*** *to complete this task. Be sure to use all the time.*

Figure 6

Mockup Messenger Chat for Generative Conditions



Note. Mockup messenger chat with a profile picture and message from the fictitious peer Mia. Students in the teaching-only condition (left), the teaching + visualization condition (middle), and the teaching + drawing condition (right) could record a voice message. Translated from German.

3.4.4 Measures

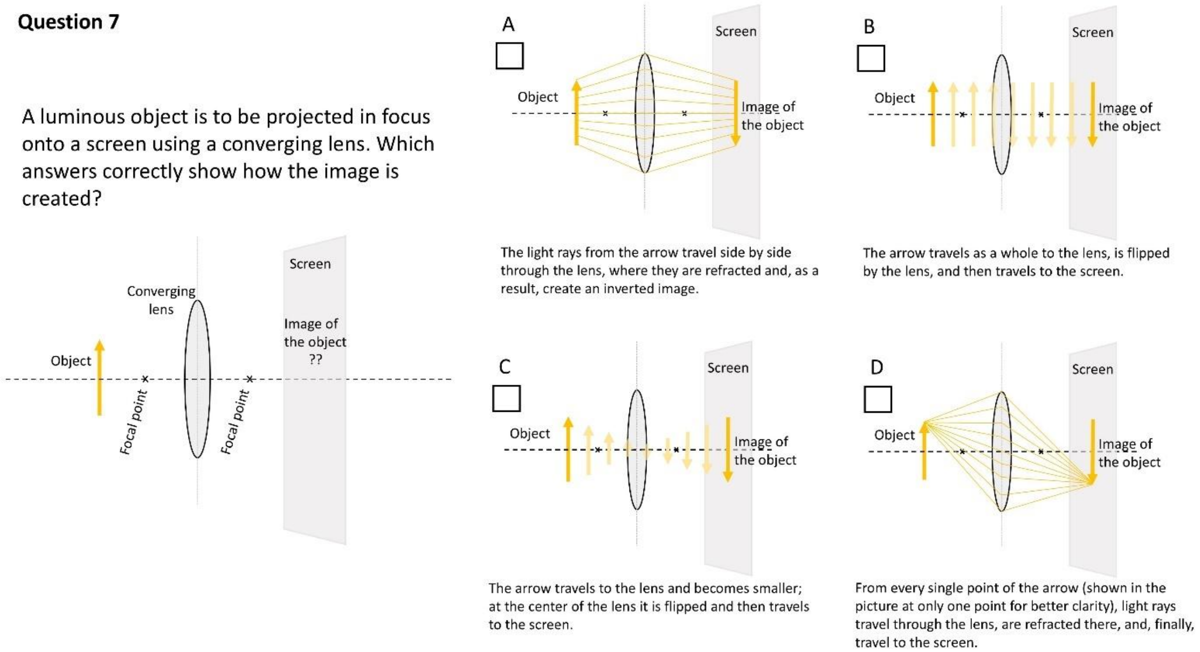
Data collection was entirely paper based. As the study was part of a larger research project, we have refrained from reporting the entire set of variables and report only those of interest for the present study. A list of all variables is provided in Appendix A.

3.4.4.1 Cognitive Learning Outcomes

To measure students' prior, immediate, and lasting cognitive knowledge, we used the validated conceptual knowledge test ROC-CI by Wörner et al. (2022). The test comprises 15 multiple-choice items that assess students' conceptual understanding of the functioning of a converging lens (e.g., "A luminous object is to be projected in focus onto a screen using a converging lens. Which answers correctly show how the image is created?") and included distractors which reflected common misconceptions (e.g., "The arrow travels as a whole to the lens, is flipped by the lens, and then travels to the screen", see Figure 7). If students selected the correct answer(s) for an item, they achieved two points, and for a partly correct answer one point (e.g., students selected only one of two correct answers; for more details see Wörner et al., 2022). In total, the students could score 30 points on the conceptual knowledge test. To minimize recognition effects, we presented the answers in the pre-, post-, and delayed test in a randomized order. Four independent raters coded 20 percent of all students' answers of the pretest. The interrater reliability was excellent ($ICC_{2,1}$ = 0.99), therefore the remaining answers were split equally between the four raters. The conceptual knowledge test indicated a satisfying level of reliability (McDonald's ω_t = 0.70).

Figure 7

Exemplary Item From the ROC-CI Conceptual Knowledge Test About Converging Lenses



Note. From Wörner et al. (2022).

3.4.4.2 Metacognitive Learning Outcomes

To evaluate how accurately students could assess their current level of understanding and make a metacognitive judgment, we measured their monitoring accuracy. To this end, we asked students to make prospective judgments about their expected performance on the pretest, the immediate posttest, and the delayed posttest: "In the following you will find 15 questions about the topic 'Imaging by converging lenses'. You can get two points per question. In total, you can score 30 points. How many points do you think you will score?" (Baars et al., 2017; Jacob et al., 2022; Prinz et al., 2018) on a scale from 0 to 30 points (see Jacob et al., 2020, 2022; Schleinschok et al., 2017, for a similar approach). We chose prospective rather than retrospective judgments to ensure that students' potential reflection processes on learning, triggered by the generative task, were captured without being influenced by students' experience with the immediate posttest (Fleming et al., 2016; see also Chua et al., 2009). This approach was consistently applied to the pretest and delayed posttest. We operationalized students' monitoring accuracy for the pretest, immediate posttest, and delayed posttest in terms of absolute differences between students' estimated performance and their actual performance without negative values (i.e., $|X_{Judgment} - X_{Performance}|$; min. score 0 points, max. score 30 points; Gutierrez et al., 2016; Gutierrez de Blume et al., 2021; Händel et al., 2020; Schraw, 2009).

3.4.4.3 Task-Specific Motivation

As an indicator of students' task-specific motivational orientations, we measured students' task interest (two items, e.g., "The task has aroused my interest", Cronbach's $\alpha = 0.86$; Jacob et al., 2021) and task enjoyment (two items, e.g., "I enjoyed doing the task", Cronbach's $\alpha = 0.84$; Jacob et al., 2021) after the learning activity. For each scale, we used a four-point Likert scale from one "I completely disagree" to four "I completely agree".

3.4.4.4 Additional Control Measures

Students' Prerequisites. To further control for differences among the experimental conditions, we measured students' interest in physics (four items; e.g., "I'm interested in what I learn in physics", McDonald's $\omega_t = 0.86$), physics work ethic (four items; e.g., "I'm paying attention in physics lessons", McDonald's $\omega_t = 0.79$), academic self-concept in physics (four items; e.g., "I even understand the most difficult tasks in physics lessons", McDonald's $\omega_t = 0.89$), and ICT interest (four items; e.g., "I like using digital devices", McDonald's $\omega_t = 0.71$). For each scale, we used a four-point Likert scale from one "I completely disagree" to four "I completely agree". All scales were adapted from Mang et al. (2018, 2019), see also Flegr et al. (2023).

Cognitive Load. After the teaching unit, students were instructed to rate their actively invested effort as an indicator of their active cognitive load (e.g., "I exerted myself for this task"; Klepsch & Seufert, 2021; see also Paas, 1992) and their passively experienced load as an indicator of their passive cognitive load (e.g., "This task was strenuous"; Klepsch & Seufert, 2021) on a Likert scale from one "not strenuous at all" to nine "very strenuous".

Affect. We used two scales to assess students' arousal and mood ("How are you feeling at the moment?", Betella & Verschure, 2016) after the teaching unit on a Likert scale from one "sleepy/ bored" to nine "wide awake/ focused" (arousal) and from one "sad/ in a bad mood" to nine "happy/ in a good mood" (mood).

Teaching Quality. To ensure a high level of implementation fidelity, we additionally assessed the teaching quality directly after the teaching unit through student evaluations. All students were asked to rate the teaching quality of the teaching unit including students' cognitive activation (six items, e.g., "In today's physics lesson on converging lenses, we were working on tasks that I had to think about very thoroughly", McDonald's $\omega_t = 0.72$; adapted from Fauth et al., 2014), student disturbances (three items, e.g., "In today's physics lesson on converging lenses, students often disturbed the lesson", McDonald's $\omega_t = 0.86$; adapted from Baumert et al., 2012), teacher monitoring (five items, e.g., "Today's physics teacher made sure that we pay attention", McDonald's $\omega_t = 0.84$; Baumert et al., 2012), and teacher support (four items, e.g., "Today's physics teacher was interested in the learning progress of each individual student", McDonald's $\omega_t = 0.82$; adapted from Mang et al., 2019). For each scale, we used a four-point Likert scale from one "I completely disagree" to four "I completely agree".

3.4.5 Procedure

The current study received approval by the Ethics Committee of the University of Tübingen and the Ministry of Education and Cultural Affairs of the State of Baden-Württemberg. Participation in the study was voluntary and we only collected data from students who had provided written consent from their legal guardians to participate. The design and procedure of the study with four experimental conditions is depicted in Figure 8.

The first author of this study visited the schools about one week before the teaching unit to personally introduce the classes to the upcoming study. This visit also aimed to assess students' demographics and prerequisites, such as prior cognitive and metacognitive knowledge, through a pretest.

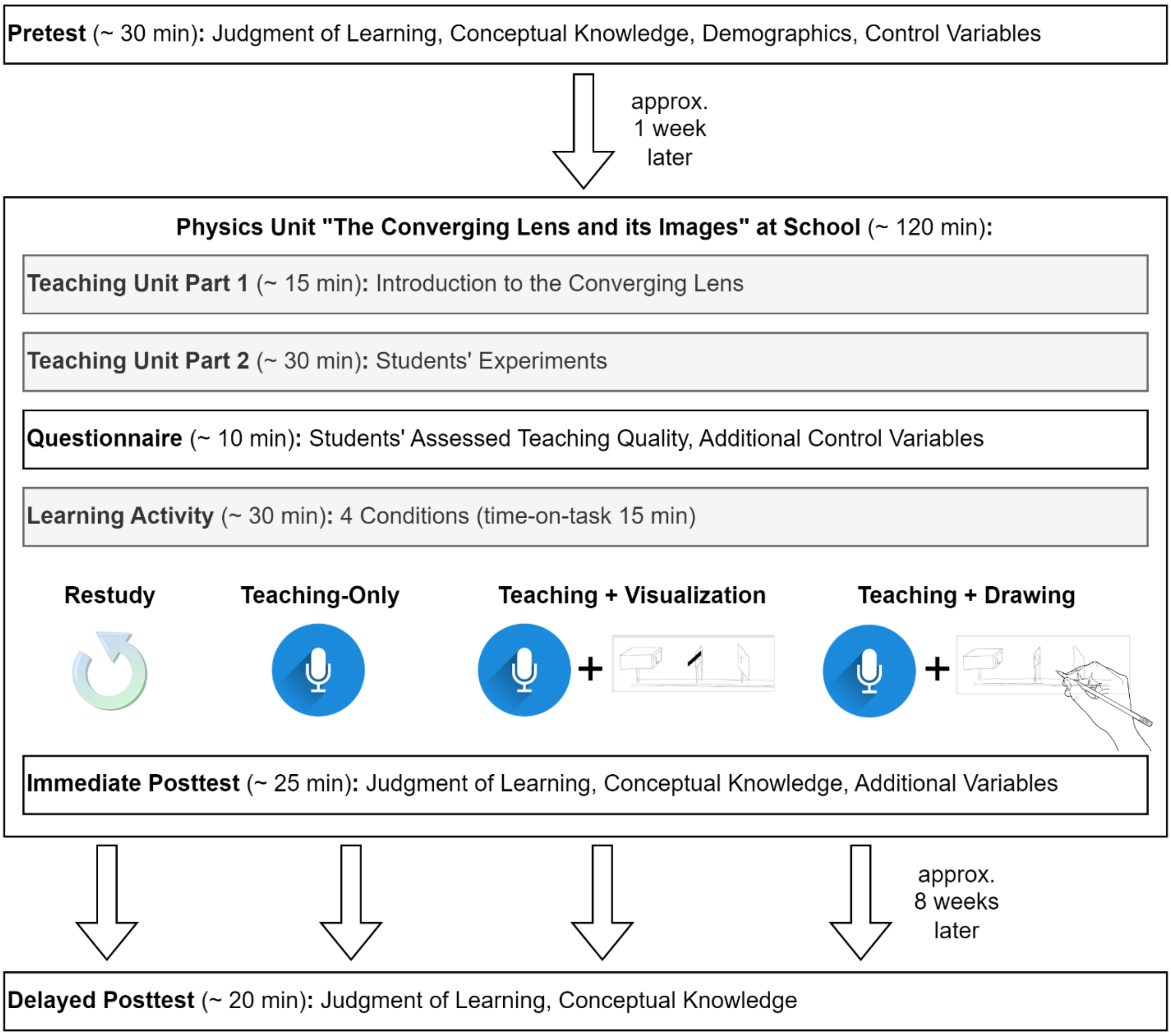
A week later, the first author—who is both a researcher and a certified physics teacher with ten years of teaching experience (hereafter referred to as "physics teacher")—personally conducted the central part of the study in all classes, including instruction and data collection. The teaching unit started with an introduction to the converging lens, at which students received a corresponding overview sheet (lasting approximately 15 minutes). Afterwards, students conducted experiments guided by an experimentation worksheet (lasting approximately 30 minutes). Communication between the students and the physics teacher proceeded in the manner of regular physics lessons in school; for example, students could ask the teacher for help if they did not understand a task. To ensure that students documented the results of the experiments correctly, the physics teacher briefly discussed these results with the students.

After the teaching unit, students answered a questionnaire in which we assessed students' cognitive load, affect, and the teaching quality. Afterwards, students were randomly assigned to one of four experimental conditions (restudy, teaching-only, teaching + visualization, teaching + drawing) within each class. During the learning activity, students either taught the learning contents to a fictitious peer, taught with a provided visualization, taught and drew, or restudied the contents in the same amount of time (15 minutes). In this learning activity phase, each student was given an individual seat. The students in the restudy condition were in one room and students in the generative conditions (teaching-only, teaching + visualization, teaching + drawing) in a second room. The students in the generative conditions wore sound-proof ear protectors to avoid potential disturbances by other students. All students had their materials available during their task. The four conditions followed this open-book format to avoid a "hidden" retrieval-practice effect and to maintain consistency in retrieval processes across all conditions (Roelle, Endres, et al., 2023; Sibley et al., 2022). That said, open-book teaching has been shown to be superior to closed-book teaching (Sibley et al., 2022).

Finally, the students completed the immediate posttest. The students' regular physics teachers were privately instructed not to further address the topic of the converging lens and its images during the following approximately eight weeks. At the end of this period, the first author revisited the classes and conducted the delayed posttest with the students.

Figure 8

Design and Procedure of the Study



3.4.6 Analysis and Coding of Students' Explanations and Drawings

The full analysis scheme for students' explanations and drawings is available at: https://doi.org/10.17605/OSF.IO/DQX3A.

3.4.6.1 Characteristics of Students' Explanations

Students in the generative conditions (i.e., teaching-only, teaching + visualization, teaching + drawing) taught the learning contents orally to their fictitious peer Mia. These explanations were transcribed, and based on prior research (e.g., Fiorella & Kuhlmann, 2020; Hoogerheide, Renkl, et al., 2019; Jacob et al., 2020, 2022), three indicators for underlying processes during teaching were coded: completeness, elaboration, and correctness.

For *completeness* of the explanations, we coded students' mentioned concepts necessary to teach the learning contents. Students could receive five points for each of the two parts of the task (see Figure 6); thus, ten points in total (e.g., one point for "the image is swapped left-right and top-bottom compared to the object"; self-generated coding scheme based on conceptual knowledge about the converging lens, Boshuizen & Schmidt, 1992; Wörner et al., 2022). Three independent raters coded 20% of the explanations. Interrater reliabilities were good to excellent (first part of the task: $ICC_{2,1}$ = 0.95, second part: $ICC_{2,1}$ = 0.88), so the remaining explanations were split equally between the three raters. The completeness score was calculated by adding both task scores.

We also counted the number of *elaborations* per explanation, including idea units such as analogies, examples, and own experiences (see also Fiorella & Kuhlmann, 2020; Jacob et al., 2020; Lachner et al., 2018). For example, the sentence "if I cover the lens ring-shaped, the image on the screen is also simply less bright than before" is an elaboration, as it was not mentioned during the learning phase. Again, three independent raters coded the number of elaborations for 20% of the explanations. Interrater reliabilities for both parts of the explanation task were excellent (first part: $ICC_{2,1}$ = 0.90, second part: $ICC_{2,1}$ = 0.94). Thus, the coding of the remaining teachings was equally split between the three raters. The elaboration score was calculated by adding both task scores.

As an indicator of the level of *correctness*, we looked at the percentage of correct idea units in the explanations (Hoogerheide, Renkl, et al., 2019). First, we counted the idea units in both parts of the teaching task; for instance, an idea unit was "the image is less bright than before". Second, we coded the number of correct idea units. Finally, we calculated the percentage of correct idea units. Three independent raters coded the percentage of correct idea units of 20% of the explanations. Interrater reliabilities for both parts of the teaching task were excellent (first part: $ICC_{2,1} = 0.99$, second part: $ICC_{2,1} = 0.95$). Thus, the coding of the remaining teachings was equally split between the three raters. The correctness score was calculated as a percentage based on the respective sum of the number of idea units and the number of correct idea units of the two parts of the task.

3.4.6.2 Characteristics of Students' Drawings

Students of the teaching + drawing condition produced drawings. Similar to the characteristics of students' explanations, three indicators of underlying processes during drawing were coded: completeness, elaboration, and correctness (see also Ainsworth & Scheiter, 2021; Fiorella & Kuhlmann, 2020; Schmidgall et al., 2019; Schwamborn et al., 2010).

For *completeness* of the drawing, we coded the concepts reflected in the students' drawings necessary to visualize the learning contents. Students could receive five points for each of the two parts of the task (see Figure 6); thus 10 points in total (e.g., one point for drawing a complete image on the screen despite a half-covered lens; self-generated coding scheme based on conceptual knowledge about the converging lens, Boshuizen & Schmidt, 1992; Wörner et al., 2022). Two independent raters coded 20% of the drawings. As interrater reliabilities were excellent (first part: $ICC_{2,1} = 0.97$, second part: $ICC_{2,1} = 1.00$), one rater coded the remaining drawings. The completeness score was calculated by adding both task scores.

We also counted the number of *elaborations* per drawing, including idea units such as analogies, examples, and own experiences (see also Fiorella & Kuhlmann, 2020; Jacob et al., 2020; Lachner et al., 2018). For example, the drawn light ray of a specific object point of the 'L' through the converging lens to its image point is an elaboration, as it was not shown during the learning phase. Again, two independent raters coded the number of elaborations for 20% of the drawings. Interrater reliabilities for both parts of the drawing task were excellent (first part: $ICC_{2,1} = 1.00$, second part: $ICC_{2,1} = 0.99$). Thus, one rater coded the remaining drawings. The elaboration score was calculated by adding both task scores.

As an indicator of the level of *correctness*, we looked at the percentage of correct knowledge in the drawings. First, we counted the idea units in both parts of the drawing task; for instance, an idea unit was the drawn upside-down image of the 'L'. Second, we coded the number of correct idea units. Finally, we calculated the percentage of correct idea units. Two independent raters coded the percent of correct thought elements for 20% of the drawings. Interrater reliabilities for both parts of the drawing task were excellent (first part: $ICC_{2,1} = 0.98$, second part: $ICC_{2,1} = 0.97$). Thus, one rater coded the remaining drawings. The correctness score was calculated as a percentage based on the respective sum of the number of idea units and the number of correct idea units of the two parts of the task.

3.4.7 Data Analyses

To test our preregistered hypotheses, we applied planned contrasts and controlled for students' prior (meta-)cognitive knowledge[4]. To account for the hierarchical structure of the data (students nested in classes nested in schools), we employed multilevel modeling and used cluster-robust standard errors. In the first contrast, we tested whether generation was more effective than restudy (i.e., restudy: -3; teaching-only: 1; teaching + visualization: 1; teaching

[4] Contrary to our preregistration, we used planned contrasts instead of ANCOVAs to more precisely test our hypotheses (Furr & Rosenthal, 2003; Rosenthal & Rosnow, 1985).

+ drawing: 1). In the second contrast, we tested whether the visualization conditions (i.e., teaching + visualization, teaching + drawing) were more beneficial than teaching-only (i.e., restudy: 0; teaching-only: -2; teaching + visualization: 1; teaching + drawing: 1). In the third contrast, we tested whether the teaching + drawing condition was better than the teaching + visualization condition (i.e., restudy: 0; teaching-only: 0; teaching + visualization: -1; teaching + drawing: 1). Additionally, we conducted mediation analyses with the contrast-coded experimental conditions as independent variable to test potential mediation effects of task-specific motivation and the characteristics of students' explanations.

Missing values naturally occur in the field study in schools across three time points. Therefore, in the preliminary and main analyses, we applied multiple imputations with 50 imputed datasets and 50 iterations.

3.4.8 Transparency and Openness

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study, and the study follows JARS (Appelbaum et al., 2018). All data, analysis code, and research materials are available at https://doi.org/10.17605/OSF.IO/DQX3A. Data were analyzed using R, version 4.4.1 (R Core Team, 2024). This study's hypotheses, design and its analysis were preregistered prospectively, before data were collected; see https://aspredicted.org/R35_NP8.

3.5 Results

We used Cohen's *d,* φ, and partial η_p^2 as effect size measures, qualifying values of d = .20, .50, .80, φ = .10, .30, .50, and η_p^2 = .01, .06, .14 as small, medium, and large effects (Cohen, 2013). We applied an alpha level of α = .05.

3.5.1 Preliminary Analysis

Analyses of the imputed data indicated that the learning activity groups did not differ concerning gender, χ^2 (6, 590) = 5.27, p = .520, φ = .09, and native language, χ^2 (6, 590) =

6.64, $p = .369$, $\varphi = .11$. A MANOVA showed no significant differences according to students' age, interest in physics, physics work ethic, academic self-concept in physics, ICT interest, prior cognitive knowledge, and prior metacognitive knowledge, $F(6,590) = 1.03$, $p = .427$, $\varphi = .21$. A further MANOVA indicated that students' cognitive load, students' affect, and teaching quality regarding the teaching unit was comparable among groups, $F(6,590) = 0.92$, $p = .645$, $\varphi = .02$. Additionally, graphical boxplot analyses did not reveal any outliers. Table 1 presents the means and standard deviations across the four conditions (i.e., restudy, teaching-only, teaching + visualization, teaching + drawing) on immediate and lasting (meta-)cognitive learning outcomes. Correlations between used variables are shown in Appendix B.

Table 1

Means and Standard Deviations for all Measurements Across Experimental Conditions

Variable	Restudy		Teaching-Only		Teaching + Visualization		Teaching + Drawing	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Students' prerequisites								
Interest in physics (1-4)	2.69	0.66	2.64	0.67	2.74	0.61	2.74	0.68
Physics work ethic (1-4)	2.90	0.53	2.91	0.64	2.99	0.53	3.00	0.54
Academic self-concept in physics (1-4)	2.64	0.66	2.70	0.73	2.83	0.67	2.83	0.68
ICT interest (1-4)	3.14	0.55	3.13	0.51	3.13	0.50	3.11	0.55
Prior knowledge								
Cognitive (0-30)	8.86	3.57	9.68	3.89	8.80	3.92	9.72	3.84
Metacognitive[a] (0-30)	8.61	5.51	8.37	5.38	8.28	5.50	8.41	5.42
Perceived ratings regarding the teaching unit								
Cognitive load								
Active cognitive load (1-9)	5.88	2.01	5.99	2.16	5.99	1.70	5.95	2.02
Passive cognitive load (1-9)	3.67	1.98	3.19	1.85	3.63	1.69	3.43	1.78
Affect								
Arousal (1-9)	4.63	2.13	4.58	2.32	4.60	2.20	4.95	2.24
Mood (1-9)	6.03	2.04	5.67	2.24	5.94	2.09	6.20	2.15
Teaching quality								
Cognitive activation (1-4)	2.63	0.50	2.65	0.52	2.68	0.47	2.69	0.50
Disturbances (1-4)	2.02	0.70	1.90	0.78	2.06	0.75	1.96	0.74
Teacher monitoring (1-4)	2.92	0.60	2.92	0.62	2.95	0.53	2.97	0.63
Teacher support (1-4)	3.15	0.64	3.18	0.62	3.34	1.46	3.25	0.59
Characteristics of students' explanations								
Completeness (0-10 points)	–	–	4.83	2.59	4.90	2.66	4.60	2.64
Elaboration (each 1 point)	–	–	4.12	5.30	3.39	4.66	4.10	7.22
Correctness (percentage)	–	–	90.04	16.16	91.62	13.57	93.53	10.30
Characteristics of students' drawings								

Variable	Restudy		Teaching-Only		Teaching + Visualization		Teaching + Drawing	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Completeness (0-10 points)	–	–	–	–	–	–	6.29	2.37
Elaboration (each 1 point)	–	–	–	–	–	–	2.59	2.83
Correctness (percentage)	–	–	–	–	–	–	87.51	13.97
Perceived ratings regarding the learning activities								
Task-specific motivation								
Task interest (1-4)	2.16	0.92	2.58	0.88	2.53	0.86	2.80	0.82
Task enjoyment (1-4)	2.17	0.92	2.45	0.88	2.42	0.88	2.60	0.87
Immediate learning outcomes								
Cognitive (0-30)	15.68	5.41	16.60	5.33	16.64	4.95	18.20	4.86
Metacognitive[a] (0-30)	5.93	4.43	5.49	4.28	5.35	4.76	5.52	4.18
Lasting learning outcomes								
Cognitive (0-30)	13.27	5.05	13.29	5.24	14.02	4.74	14.24	5.41
Metacognitive[a] (0-30)	6.53	4.98	5.71	4.48	5.64	4.40	5.10	4.05

Note. The data in this table are based on the raw data.

[a]Metacognitive learning outcomes are measured by monitoring accuracy, defined by the absolute difference between students' predicted and actual performance without negative values.

3.5.2 Cognitive Learning Outcomes

In line with the generation hypothesis (H1a), contrast analyses revealed that students who taught the learning contents to a fictitious peer (i.e., teaching-only, teaching + visualization, teaching + drawing) outperformed students who restudied the learning contents in the immediate posttest (β = 0.06, *p* = .008, small effect). To explore whether task-specific motivational mechanisms underlie this significant generation effect, as contrasted to restudying the learning contents, we conducted a mediation analysis. Learning activity was the predictor (generation contrast: -3 = restudy, 1 = teaching-only, 1 = teaching + visualization, 1 = teaching + drawing), task interest and task enjoyment were the mediators, and students' immediate cognitive learning outcome was the dependent variable. Results indicated a significant indirect effect via task interest of $a_1 \times b_1 = 0.06$, *p* = .012. Thus, students who taught reported higher interest in the learning activity which resulted in higher learning outcomes. Task enjoyment was not a significant mediator (see Figure 9 for the full mediation model).

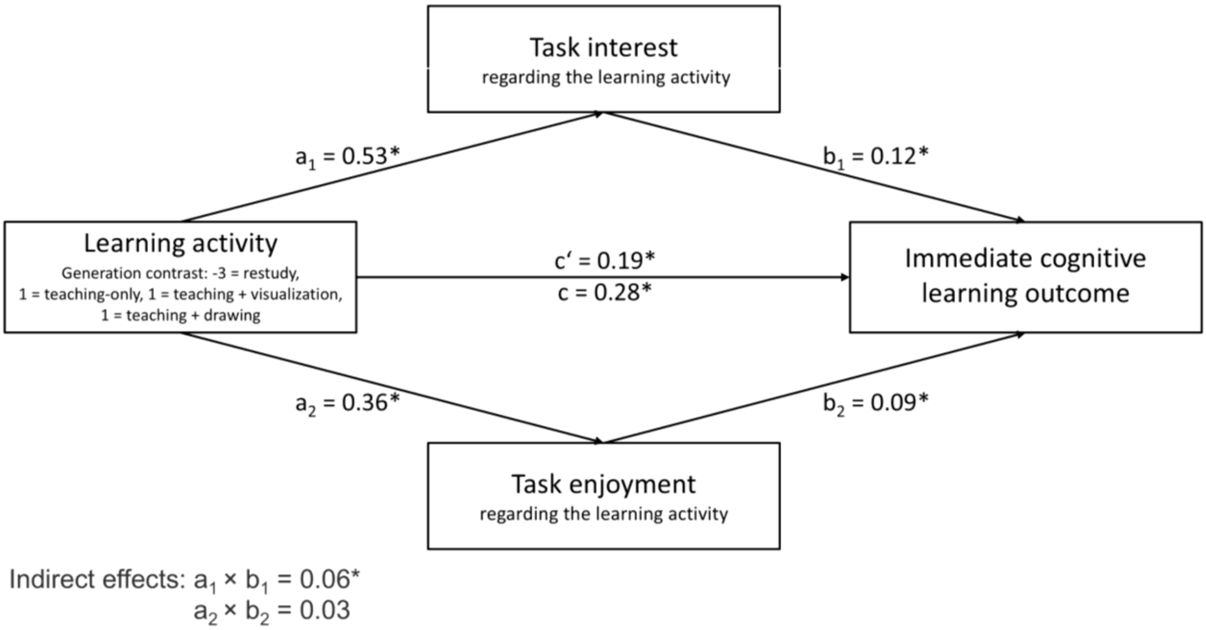
Consistent with our visualization hypothesis (H2a), the second contrast was also significant. Thus, adding a visualization component (teaching + visualization, teaching + drawing) led to higher learning outcomes than teaching-only in the immediate posttest (β = 0.06, *p* = .045, small effect). The examination of the descriptive values for immediate cognitive learning outcomes (see Table 1) suggests that this visualization effect is primarily attributable to the teaching + drawing condition. These considerations made it especially important to examine whether teaching + drawing was more beneficial than teaching + visualization.

In fact, and in line with our drawing hypothesis (H3a), teaching + drawing was more beneficial than teaching + visualization (β = 0.11, *p* = .037, small effect). We explored the mechanisms underlying this significant teaching and drawing effect, as contrasted to teaching with a provided visualization of the learning contents**.** Accordingly, we analyzed the correlations between task interest, identified as mediator of our generation contrast, the characteristics of students' explanations (completeness, elaboration, correctness), and immediate cognitive learning outcome (see Appendix B). These explorative analyses suggest that students in the teaching + drawing condition showed higher immediate cognitive learning outcomes than students in the teaching + visualization condition, as the combination of teaching and drawing triggered higher task interest concerning the learning activity, which resulted in more complete explanations (indicated by the points of completeness). We explored this mediation assumption by conducting a serial mediation analysis (task interest → completeness). Learning activity was the predictor (drawing contrast: 0 = restudy, 0 = teaching-only, -1 = teaching + visualization, 1 = teaching + drawing), task interest and the points of completeness were the serial mediators, and students' immediate cognitive learning outcome was the dependent variable. Results indicated a significant indirect effect via task interest and completeness of $a_1 \times d_{21} \times b_2 = 0.03$, $p = .029$ (see Figure 10 for the full mediation model). Thus, students who taught and drew reported higher interest in the learning activity, which led to more complete explanations and resulted in higher learning outcomes. No other significant mediators were identified (for detailed results, see Table 3).

Next, we analyzed potential differences between the conditions regarding students' lasting cognitive learning outcomes as assessed in the delayed posttest, conducted eight weeks after the intervention. The planned contrasts did not reveal any significant differences. Specifically, the comparison between the restudy control condition and the generative conditions (i.e., teaching-only, teaching + visualization, teaching + drawing) resulted in a non-significant effect ($\beta = 0.03$, $p = .198$). Similarly, contrasting teaching-only with the visualization conditions (i.e., teaching + visualization, teaching + drawing) showed no significant differences ($\beta = 0.06$, $p = .077$). Finally, the comparison between teaching + visualization and teaching + drawing yielded no significant effect ($\beta = -0.01$, $p = .892$). These findings suggest that neither generative activities nor the inclusion of visualization or drawing contributed more to lasting cognitive learning outcomes compared to the respective contrasted conditions (for more details, see Table 2).

Figure 9

Mediation Analysis in Terms of Task Interest and Task Enjoyment



Note. *p < .050. Regression coefficients are standardized.

Table 2

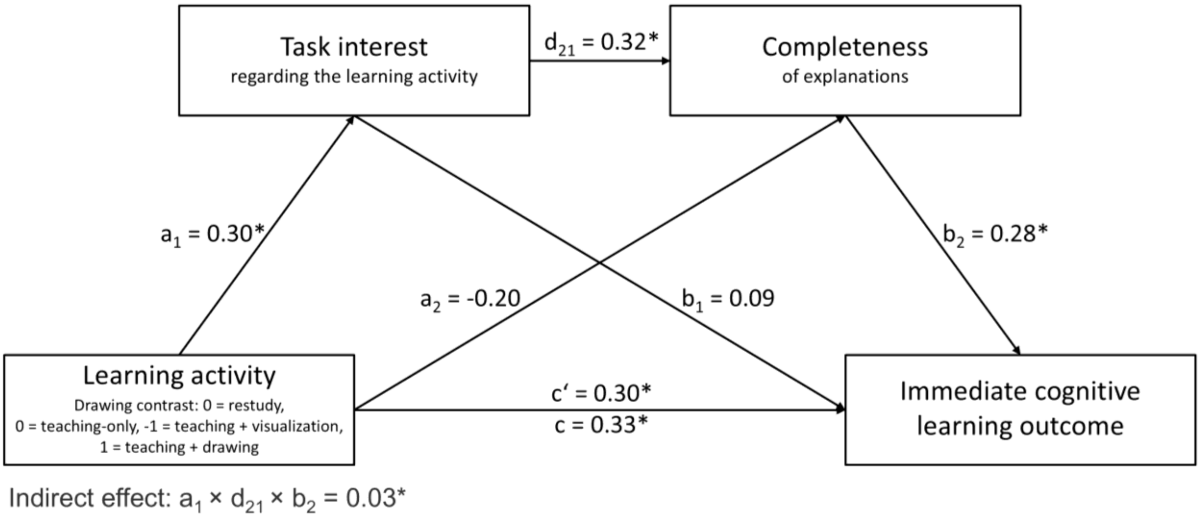
Summary of the Contrast Analyses on (Meta-) Cognitive Learning Outcomes

Variable	β	*SE*	*t*	*p*
Cognitive learning outcomes				
Immediate				
Generation contrast[a]	**0.06**	**0.02**	**2.65**	**.008**
Visualization contrast[b]	**0.06**	**0.03**	**2.01**	**.045**
Drawing contrast[c]	**0.11**	**0.05**	**2.09**	**.037**
Prior cognitive knowledge	**0.32**	**0.04**	**7.70**	**<.001**
Lasting				
Generation contrast[a]	0.03	0.02	1.29	.198
Visualization contrast[b]	0.06	0.03	1.77	.077
Drawing contrast[c]	-0.01	0.05	-0.14	.892
Prior cognitive knowledge	**0.43**	**0.04**	**10.31**	**<.001**
Metacognitive learning outcomes				
Immediate				
Generation contrast[a]	-0.03	0.02	-1.14	.255
Visualization contrast[b]	0.00	0.03	-0.12	.904
Drawing contrast[c]	0.02	0.06	0.30	.763
Prior metacognitive knowledge	0.04	0.05	0.76	.447
Lasting				
Generation contrast[a]	-0.05	0.03	-1.92	.056
Visualization contrast[b]	-0.03	0.04	-0.93	.351
Drawing contrast[c]	-0.06	0.06	-1.11	.269
Prior metacognitive knowledge	**0.11**	**0.05**	**2.10**	**.036**

Note. Significant results are highlighted in bold letters; $p < .050$.
[a] -3 = restudy, 1 = teaching-only, 1 = teaching + visualization, 1 = teaching + drawing. [b] 0 = restudy, -2 = teaching-only, 1 = teaching + visualization, 1 = teaching + drawing. [c] 0 = restudy, 0 = teaching-only, -1 = teaching + visualization, 1 = teaching + drawing.

Figure 10

Serial Mediation Analysis in Terms of Task Interest and Completeness of Students' Explanations



Note. *p < .050. Regression coefficients are standardized.

Table 3

Summary of the Mediation Analyses With Task Interest and Characteristics of Students' Explanations Regarding Immediate Cognitive Learning Outcome

Variable	β	*SE*	*t*	*p*
Task interest as mediator				
Effect of drawing contrast[a] on task interest (a)	**0.30**	**0.11**	**2.78**	**.005**
Effect of task interest on learning outcome (b)	**0.17**	**0.06**	**2.99**	**.003**
Direct effect (c')	**0.25**	**0.11**	**2.27**	**.023**
Indirect effect (a*b)	**0.05**	**0.03**	**2.04**	**.042**
Total effect (c)	**0.30**	**0.11**	**2.74**	**.006**
Characteristics of students' explanations as mediator				
Completeness				
Effect of drawing contrast[a] on completeness (a)	-0.10	0.12	-0.87	.387
Effect of completeness on learning outcome (b)	**0.30**	**0.05**	**5.49**	**<.001**
Direct effect (c')	**0.33**	**0.10**	**3.17**	**.002**
Indirect effect (a*b)	-0.03	0.04	-0.86	.393
Total effect (c)	**0.30**	**0.11**	**2.74**	**.006**
Elaboration				
Effect of drawing contrast[a] on elaboration (a)	0.12	0.12	0.97	.333
Effect of elaboration on learning outcome (b)	**0.16**	**0.05**	**3.11**	**.002**
Direct effect (c')	**0.28**	**0.11**	**2.61**	**.009**
Indirect effect (a*b)	0.02	0.02	0.92	.356
Total effect (c)	**0.30**	**0.11**	**2.74**	**.006**
Correctness				
Effect of drawing contrast[a] on correctness (a)	0.14	0.10	1.31	.191
Effect of correctness on learning outcome (b)	**0.27**	**0.06**	**4.22**	**<.001**
Direct effect (c')	**0.26**	**0.11**	**2.48**	**.013**
Indirect effect (a*b)	0.04	0.03	1.25	.211
Total effect (c)	**0.30**	**0.11**	**2.74**	**.006**

Note. Significant results are highlighted in bold letters; $p < .050$.
[a] 0 = restudy, 0 = teaching-only, -1 = teaching + visualization, 1 = teaching + drawing.

3.5.3 Metacognitive Learning Outcomes

For students' immediate monitoring accuracy, contrast analyses revealed no significant differences between the conditions. Specifically, no effect was found when comparing the restudy control condition with the generative conditions (i.e., teaching-only, teaching + visualization, teaching + drawing; $\beta = -0.03$, $p = .255$), which was contrary to our generation hypothesis (H1b). Similarly, and not in line with our visualization hypothesis (H2b) or our drawing hypothesis (H3b), no significant effect was observed for the comparison between teaching-only and the visualization conditions (i.e., teaching + visualization, teaching + drawing; $\beta = 0.00$, $p = .904$) or between teaching + visualization and teaching + drawing ($\beta = 0.02$, $p = .763$).

For lasting monitoring accuracy, the results similarly showed no significant differences. The comparison between the restudy control condition and the generative conditions (i.e., teaching-only, teaching + visualization, teaching + drawing) yielded a non-significant effect ($\beta = -0.05$, $p = .056$). Likewise, no significant differences were found between teaching-only and the visualization conditions (i.e., teaching + visualization, teaching + drawing; $\beta = -0.03$, $p = .351$) or between teaching + visualization and teaching + drawing ($\beta = -0.06$, $p = .269$). These findings indicate that neither generative activities nor the addition of visualization or drawing enhanced students' accuracy in their judgments of learning after an eight-week delay when compared to the students in the respective contrasted conditions (for more details, see Table 2).

3.6 Discussion

In this large-scale experimental classroom study with secondary school students, we aimed at investigating whether the combination of students' non-interactive teaching and drawing enhances their (meta-)cognitive learning within an authentic inquiry-based physics classroom environment regarding both immediate and lasting learning (after eight weeks). We also aimed at examining whether combining non-interactive teaching and drawing was more effective due to the visualization of the learning contents or due to the active generation of drawings. Additionally, we explored students' task-specific motivation and characteristics of students' explanations as underlying mechanisms of a potential effect.

For immediate cognitive learning outcomes, in line with our generation hypothesis (H1a), our results demonstrated that students who engaged in non-interactive teaching (i.e., teaching-only, teaching + visualization, teaching + drawing) outperformed students who restudied the contents. Our findings highlight that non-interactive teaching is an effective learning strategy to boost inquiry-based education in school. These results align with the limited existing evidence on non-interactive teaching among students in a school context, such as the study by Hoogerheide, Visee, et al. (2019). However, the authors (Hoogerheide, Visee, et al., 2019) did not implement their study in authentic inquiry-based lessons, used self-contained expository learning materials, and focused on primary students. Moreover, their students engaged in generative learning at home rather than in class. In contrast, we demonstrated that the non-interactive teaching effect (Kobayashi, 2024; Lachner et al., 2021) not only replicates but also generalizes within authentic inquiry-based natural science lessons with secondary school students.

Additionally, our explorative mediation analysis revealed that the generation effect was explained by increased task interest regarding the generative conditions (teaching-only, teaching + visualization, teaching + drawing) contrasted to the restudy control condition. Similar results can be found in prior research of non-interactive teaching that also demonstrated that the task interest may be a crucial factor for students' learning (Jacob et al., 2021).

As expected, based on our visualization hypothesis (H2a), adding a visualization component (i.e., teaching + visualization, teaching + drawing) significantly enhanced the effectiveness of non-interactive teaching contrasted to teaching-only. This suggests that students' teaching with a visualization component helped them construct more meaningful representations of the learning contents. Our results align with those of Fiorella (2023a), who demonstrated that students using visualizations (provided visualization, drawing) outperformed those using words (provided words, generating words) during non-interactive teaching. However, Fiorella's study (2023a) is the only one to have specifically investigated this effect, and its findings were limited to a transfer test. Additionally, the laboratory study relied on retrieval processes through a closed-book design, where teaching itself served as a retrieval activity. Our results extend this visualization effect to an authentic, inquiry-based learning context in school, underscoring its relevance in real-world educational contexts. Additionally, the descriptive analysis of immediate cognitive learning outcomes indicated that the observed visualization effect can largely be attributed to the teaching + drawing condition. This raises the question of whether teaching + drawing offers unique benefits over teaching + visualization alone. These findings make it particularly important to examine whether drawing as a generative learning strategy contributes more significantly to students' cognitive gains than simply providing visualizations during non-interactive teaching.

In fact, and in line with our drawing hypothesis (H3a), students who taught and drew outperformed students who taught with a provided visualization. As opposed to Fiorella (2023a), our results highlight that students who taught and generated a drawing constructed more coherent mental representations and consequently higher learning outcomes contrary to teaching with a provided visualization of the learning contents. Therefore, the combination of teaching and drawing is effective not only due to its generative affordances, but rather due to the process of generating meaningful visuospatial representations of the learning contents through the act of drawing. These findings provide an important contribution to the field, as we were, to our knowledge, the first to systematically examine the relative effectiveness of different visualization formats during non-interactive teaching in an authentic inquiry-based learning context in school (Ainsworth & Scheiter, 2021; Fiorella & Zhang, 2018).

Explorative analyses of underlying mechanisms suggested that the effect of combining non-interactive teaching and drawing was explained by increased task interest and increased completeness of the generated explanations. Apparently, drawing additionally increased students' task interest and supported students in generating more complete explanations. These findings demonstrate both underlying cognitive and task-specific motivational processes. Moreover, drawing has been shown to facilitate teaching. The results of our present study are consistent with predictions from generative learning theory (Fiorella & Mayer, 2015, 2016) and cognitive-affective theory of learning with media (Moreno & Mayer, 2007). For instance, Pintrich (2003) presents the importance of students' motivational factors in facilitating cognition in his comprehensive review of motivational and emotional components related to classroom factors and school performance of students. Students' motivation is a key factor in getting students to initiate and maintain meaningful generative processes (Fiorella & Mayer, 2016), thus resulting in increased cognitive engagement (Moreno & Mayer, 2007; Pintrich, 2003). Thus, both cognitive but also motivational processes seem to be important underlying mechanisms of generative learning strategies and therefore should be considered and analyzed in future studies in more detail.

Our findings on the effectiveness of combined non-interactive teaching and drawing as generative learning strategies invite reflection on why drawing might have been more successful in the present study compared to some previous research. Notably, our drawing activity was embedded into curriculum-aligned authentic lessons, framed by a story featuring the fictitious peer Mia, in which students used drawing as a means to explain concepts to Mia, and conducted in students' familiar school context. This likely increased its relevance and fostered engagement compared to prior studies (e.g., Fiorella, 2023a; Fiorella & Jaeger, 2023). Unlike Ploetzner and Fillisch's (2017) study, where students observed more abstract and complex dynamic system animations and then generated corresponding drawings, our design incorporated hands-on experiments, possibly making it easier for students to represent the learning contents through drawing (see also McNeil & Uttal, 2009). Moreover, these hands-on activities likely enabled students to experience the meaning of the learning contents more directly and deeply, enhancing their conceptual understanding (for related evidence, see Sarama & Clements, 2009). Another important difference lies in the scaffolding provided during the drawing process. Unlike prior studies (e.g., with university students: Fiorella & Jaeger, 2023; with primary students: Van Essen & Hamaker, 1990), our study provided seventh and eighth-grade students with template pictures to draw on. This scaffolding likely reduced cognitive load in this age group and enhanced the quality of the generated drawings (see also Ainsworth & Scheiter, 2021; Fiorella & Zhang, 2018). This aligns with Fiorella's generative sense-making framework (2023b), suggesting that the characteristics of the students and learning materials can influence the effectiveness of drawing activities. Importantly, we combined drawing with the verbal generative activity of non-interactive teaching, both of which students used to explain the learning contents to their fictitious peer, likely strengthening generative processes and contributing to the overall effectiveness of the drawing activity. This tailored combination of non-interactive teaching and drawing, curriculum alignment, authentic lessons, story framing, hands-on experiments, and template-based scaffolding may have been particularly effective in supporting secondary students' immediate learning by balancing cognitive demands with structured support.

For immediate metacognitive learning outcomes, contrary to our hypotheses (H1b, H2b, H3b), none of the effects were significant. This contradicts prior research that showed beneficial effects of generative learning activities regarding students' monitoring accuracy (Fiorella & Jaeger, 2023; Fukaya, 2013). However, Fiorella and Jaeger (2023) distinguished between teaching and test monitoring judgments, which may have provided a clearer framework for students to evaluate their learning more accurately. This distinction could suggest that different types of monitoring judgments require targeted instructional strategies to be effective, which might explain the differences in monitoring accuracy observed in their study compared to ours (see also Schraw, 1994). Furthermore, in Fukaya's study (2013), an experimenter gave students an example to make the instructions regarding the monitoring judgment clear, before students judged their expected performance on the test. Probably, our students would have benefited from more guidance through the generative processes, provided by additional instructional clarity through examples (Fukaya, 2013), an explanation of the importance of judgments of learning (Callender et al., 2016), structured feedback on performance (Callender et al., 2016) and monitoring accuracy (Morphew, 2021; H. Wang et al., 2023), or support for fostering students' self-regulation (Van Meter, 2001). Future research should address these issues systematically. Another possible explanation for the null findings regarding monitoring accuracy could be the age of the students. For secondary students, our generative activities may have been too demanding (Brod, 2021; Fiorella, 2023b), which could have led to difficulties in accurately assessing their own performance compared to university students (e.g., Fiorella & Jaeger, 2023; Fukaya, 2013). This challenge might be particularly pronounced when generative tasks are combined such as teaching and drawing, where younger students may struggle more with monitoring their learning (see also Brod, 2021; Van Essen & Hamaker, 1990; Van Meter, 2001; Van Meter et al., 2006). Future research should investigate whether and how these challenges could be mitigated and monitoring accuracy improved among younger students, for instance by providing additional instructional support (Brod, 2021; Van Meter, 2001; Wu & Rau, 2019).

Finally, we did not find evidence of lasting learning effects concerning (meta-)cognitive outcomes, which is contrary to our expectations based on generative learning theory. Generally, this theory suggests that students who engage in generative activities should construct elaborated and consolidated mental representations of the learning contents, leading to lasting learning (Fiorella, 2023b; Fiorella & Mayer, 2016). However, research demonstrating lasting learning outcomes within non-interactive teaching environments remains scarce, as most prior studies have concentrated on immediate effects (exception with a one-week-delay, e.g., Fiorella & Mayer, 2013). Importantly, this highlights a broader gap in the literature: we know relatively little about the long-term effectiveness of generative learning strategies beyond relatively short-term evaluations like a one-week delayed posttest. Unlike Fiorella and Mayer's (2013) findings on non-interactive teaching, which reported beneficial results after a one-week delay, our study did not reveal significant lasting learning after an eight-week period. This suggests that the effects of non-interactive teaching may only persist for a short time and may diminish over longer periods such as the eight weeks in our study. It is also possible that, due to the open-book design of our study, retrieval processes potentially essential for lasting learning were not performed. Future research should track the development of knowledge after generative interventions more closely through multiple measurement points. Additionally, exploring strategies to bolster non-interactive teaching, such as incorporating consolidation techniques like retrieval practice (T. Richter et al., 2022; Roelle, Schweppe, Endres, Lachner, von Aufschnaiter, et al., 2022; Roelle, Endres, et al., 2023), may prove valuable in fostering lasting learning. To date, retrieval practice is the only approach that has been sufficiently tested and shown to produce lasting learning benefits over significant delays (see Rowland, 2014). Combining generative learning strategies with retrieval practice could further enhance lasting learning (see also Endres et al., 2024; Roelle, Schweppe, Endres, Lachner, von Aufschnaiter, et al., 2022; Roelle, Endres, et al., 2023). Future research could additionally compare closed-book teaching to retrieval practice (e.g., free recall) or to more guided forms of retrieval-based teaching, to clarify the conditions under which these approaches best support lasting learning outcomes.

3.6.1 Limitations and Future Directions

To our knowledge, we were the first to investigate whether the combination of students' non-interactive teaching and drawing is effective in real-world physics lessons in school. However, it is important to note that we focused specifically on inquiry learning within the topic of "the converging lens and its images" and conducted our study with seventh and eighth-grade school students. Future research should examine whether and how our results can be extended to other (physics) topics and subjects in school. Additionally, given that previous research has demonstrated the significance of student age in generative learning activities (Brod, 2020), it would be valuable for future studies to investigate the effects across different age groups.

While our study was conducted within the context of authentic lessons and aligned with the curriculum for physics education, it deviated in certain respects from typical classroom activities, which may have influenced the results. To ensure consistency and control, the entire teaching unit on the converging lens and its images, as well as the study-related components such as pre-, post-, and delayed tests, were implemented in all classes by the first author personally, who is both a researcher and an experienced physics teacher. However, this approach differs from regular teaching practices, where classroom teachers typically work with the same students over an extended period and are familiar, for example, with their individual learning needs and classroom dynamics. Such differences might have influenced the natural flow of classroom interactions.

Additionally, the study's assessment components diverged from typical classroom performance evaluations. Specifically, the immediate, post-, and delayed tests were no-stakes, meaning they had no consequences for the students, which is common in educational research to measure learning outcomes without introducing the confounding effects of examination pressure (e.g., Hoogerheide, Visee, et al., 2019; Lachner et al., 2020). This approach contrasts with the standard school practice of primarily using assessments with consequences for the students, such as grades or class standing. Prior research suggests that assessments with consequences for students can influence their motivation and performance (Finn, 2015; Mislevy, 1995; Pintrich & De Groot, 1990; Wolf & Smith, 1995). Although this factor was consistent across all three no-stake assessments in our study, it may have affected students' engagement and, consequently, the outcomes.

Another key deviation was that regular physics teachers were explicitly instructed not to address the topic of the converging lens and its images during the eight weeks between the immediate and delayed posttest. While this procedure was critical in isolating the effects of the interventions, it does not reflect typical teaching practice, where topics are often revisited and reinforced periodically. Furthermore, it is unclear whether some students independently looked at the topic again, for instance at home, introducing a variable that was not controlled.

Despite these deviations, the study design was necessary to systematically investigate the effects of non-interactive teaching under controlled conditions. Future research could explore ways to integrate such interventions more seamlessly into standard classroom routines, reducing deviations from typical teaching practices while maintaining methodological rigor. This could include closer collaboration with classroom teachers to implement interventions in a way that reflects authentic teaching and learning dynamics more closely.

A further potential limitation of our study is the repeated use of the same conceptual knowledge test across the pretest, immediate posttest, and delayed posttest. Aside from the fact that prior research has shown that repeated testing can enhance learning and memory retention (Rowland, 2014; Yang et al., 2021), using the same test at all three time points may have also introduced recognition effects. We decided to use the same test across the three time points because it is the only validated conceptual knowledge test available in this specific domain of converging lenses (ROC-CI; Wörner et al., 2022), ensuring reliable and consistent measurement of students' learning outcomes across conditions. If we had added or modified items, the variance structure of the test would have changed, complicating the interpretation of results across the three measurement points. To mitigate this risk, we randomized the order of the answer options in each test to reduce the likelihood that students relied solely on memory rather than genuinely recalling the concepts. Still, we cannot entirely rule out the possibility that students answered based on recognition processes rather than true conceptual recall. This limitation of using the same conceptual knowledge test three times may have impacted the ability of the test to accurately capture the actual learning effects after eight weeks, potentially contributing to the lack of a delayed effect. Furthermore, the repeated use of the same test items may have also influenced the metacognitive results. Although students were not explicitly informed that the same questions would reappear, it is possible that they anticipated this, which could have affected their prospective judgments of learning. As a result, the monitoring accuracy measurements might reflect an influence of familiarity with the test questions rather than solely students' actual knowledge or understanding. Although the repeated use of the same test may have influenced our (meta-)cognitive results in general, this limitation applied equally across all conditions, ensuring comparability. Future studies could address this issue by integrating measures that differ between the pretest, the immediate posttest, and the delayed posttest, or by employing alternative or equivalent test formats for repeated assessments.

A further limitation of this study is the lack of an assessment of knowledge transfer, which has been considered an important outcome in related research. For instance, Jacob et al. (2020) demonstrated that teaching can promote transfer outcomes, even when no significant differences in conceptual understanding are observed. In our study, the focus was placed exclusively on conceptual knowledge, as this represented the primary objective of the teaching unit and the accompanying materials (see Flegr et al., 2023; Wörner et al., 2022). However, this specific focus constrains the extent to which our findings can inform the broader advantages associated with non-interactive teaching. To provide a more comprehensive understanding of the effects of non-interactive teaching, future studies should examine transfer outcomes alongside conceptual knowledge.

In this study, we investigated whether we find an overall effect of combining students' teaching and drawing in real inquiry-based physics classes. Although the answer to this overarching question is positive and absolutely essential for further designs of real learning environments in schools, the fit of the study phase and the generative learning activities to learner characteristics such as prior knowledge or cognitive abilities, and therefore also to the performance level of the students should be critically assessed. For example, the learning contents and materials for the generative learning activity should be appropriate to students' knowledge, not require knowledge which the students do not have yet, and not redundantly provide knowledge which students already know very well (Fiorella, 2023b; Kalyuga, 2014). Further research should address learners' individual differences regarding appropriate learning contents and materials. Furthermore, future studies should explore potential adaptive arrangements for implementing generative learning activities in classrooms with heterogeneous student populations.

Another possible limitation is that the effect sizes of our study are relatively modest. This may be due to the fact that students only engaged in drawing and/or teaching once and for a relatively brief 15-minute period, which could also explain why the benefits of generative learning did not last across an 8-week period. Importantly, however, the way we implemented learning by non-interactive teaching is in line with almost every prior study, and relative to most prior non-interactive teaching research, this study already provided students with more time for generative learning (i.e., 15 minutes; Hoogerheide et al., 2016; Lachner et al., 2022). Furthermore, considering the brief intervention within a real-world school setting and the fact that small to medium effects are typical in this research area, as shown in previous meta-analyses (e.g., Lachner et al., 2022; Ribosa & Duran, 2022), these effects are substantial. Building on these findings, we argue that our results highlight an important avenue for future research, namely to investigate how the effectiveness of non-interactive teaching can be increased and made to last. One way to address this could involve systematically manipulating both the dosage of the generative activity and the retention interval. Such investigations might reveal that extending the delay beyond what was used in prior studies requires a corresponding increase in the time-on-task for generative learning activities. Another way might be to adjust the timing of the generative activity by ensuring that students can study again after teaching, which would allow them to remedy any knowledge gaps they detected during the teaching phase (cf. Lachner et al., 2020). Moreover, teaching multiple times rather than once—perhaps across multiple study sessions (cf., spacing effect; Carpenter et al., 2012a; Ebersbach et al., 2022)—might boost students' metacognition and learning and possibly reduce how much students forget over time.

While the effect sizes observed in our study are modest, they hold practical significance in real-world educational contexts. Scalable and cost-effective interventions, even with small effects, can have substantial cumulative impacts when implemented across larger populations (see Kraft, 2020). In our study, the curriculum-aligned implementation of non-interactive teaching and drawing within an inquiry-based authentic learning environment represents a practical and realistic approach to integrating generative learning strategies into everyday classroom practices. This alignment with existing educational structures not only enhances the feasibility of scaling but also underscores the potential for these interventions to contribute meaningfully to science education and, more broadly, to be extended to STEM education.

To facilitate implementation, teachers could integrate non-interactive teaching and drawing activities at the end of inquiry phases to help students deepen their understanding of the learning contents. For example, teachers might focus on a small number of clearly defined learning objectives during their lessons and incorporate hands-on experiments or concrete materials. Afterward, they could provide a chat message to prompt students' verbal explanations and drawings—for instance, as a homework assignment (Hoogerheide, Visee, et al., 2019). These activities require minimal technological resources and can be embedded in existing lesson structures. Future research should examine the further scalability and long-term feasibility of embedding generative activities like non-interactive teaching and drawing into everyday classroom practices.

To support broader implementation, school leaders, teacher educators, or curriculum developers could advance uptake by incorporating non-interactive teaching and drawing activities into instructional and professional development or aligning them with inquiry-based science curricula. Moreover, educational policymakers could embed these strategies into relevant instructional design guidelines and curriculum frameworks. This would help ensure that generative activities become integrated components of a coherent instructional design, informing educational practice at scale.

3.7 Conclusion

Our findings demonstrate that combining learning by non-interactive teaching and drawing is a beneficial approach to supporting students' cognitive learning within an authentic inquiry-based physics classroom environment. Moreover, our findings suggest that the combination of students' teaching and drawing is not only effective because of its generative affordances but rather because of the process of generating meaningful visuospatial representations of the learning contents through the act of drawing. Due to differences in task interest between teaching and drawing versus teaching with a provided visualization of the learning contents, generating teachings and drawings may best promote students' motivation and cognitive learning. Future research should examine how to further enhance the benefits of generative activities to ensure these gains last over time.

4

STUDY 2

DOES DISTRIBUTING NON-INTERACTIVE TEACHING CONTRIBUTE TO LEARNING? STUDENTS' ACADEMIC SELF-CONCEPT AND WORK ETHIC MATTER

Russ, H., Sibley, L., Flegr, S., Kuhn, J., Hoogerheide, V., Scheiter, K., & Lachner, A. (2025). Does distributing non-interactive teaching contribute to learning? Students' academic self-concept and work ethic matter. *Learning and Individual Differences,* 120, 102687. https://doi.org/10.1016/j.lindif.2025.102687

4 Study 2: Does Distributing Non-Interactive Teaching Contribute to Learning? Students' Academic Self-Concept and Work Ethic Matter

Abstract

Explaining learning contents to a fictitious peer (i.e., non-interactive teaching) improves learning, yet this effect is modest, heterogeneous, and likely influenced by individual differences. We examined whether the effectiveness of non-interactive teaching could be increased by incorporating drawing or distributing teaching. We realized a 3 × 2 field experimental design (*N* = 317), crossing the factors learning activity (restudy, teaching-only, teaching + drawing) and timing (after the study phase or distributed three times throughout the study phase). Overall, teaching resulted in better immediate conceptual knowledge than restudying, mediated by the level of completeness. This teaching effect was most pronounced in the after-study condition. However, drawing did not enhance conceptual knowledge. Students who taught underestimated their immediate knowledge. No lasting effects were observed. Students with higher academic self-concept or work ethic benefited more from teaching, highlighting the moderating role of inter-individual differences for instructional interventions.

Educational Relevance and Implications Statement

This classroom study demonstrates that non-interactive teaching is an effective instructional method in secondary school physics education. The findings highlight the importance of considering students' individual differences, such as academic self-concept or work ethic, when designing such learning activities. These insights emphasize the need for adapted and differentiated approaches that can better account for individual differences, ensuring that non-interactive teaching can be effective across diverse student populations.

Keywords

Generative learning, learning by teaching, drawing, academic self-concept, work ethic

4.1 Introduction

Learning by teaching is widely recognized as an effective way to enhance learning (Fiorella & Mayer, 2016; Pi et al., 2021). Recently, it has been shown that teaching previously learned contents even to a fictitious non-present peer (cf. non-interactive teaching), is a powerful generative activity (Fiorella & Mayer, 2013; Hoogerheide et al., 2014; Lachner et al., 2022). Non-interactive teaching may help students construct a meaningful mental representation of the learning contents, contributing to learning and metacomprehension (Fiorella, 2023b; Fiorella & Mayer, 2016; Wittrock, 1989). However, previous research indicated that the effectiveness of non-interactive teaching is relatively modest (Kobayashi, 2024; Lachner et al., 2021) and that there is large heterogeneity among the findings (Hoogerheide, Visee, et al., 2019; Jacob et al., 2022). These findings suggest that the effectiveness of non-interactive teaching may depend on students' prerequisites (e.g., Jacob et al., 2022) and that optimizing its potential may require modifications, such as combining it with other strategies.

To address these gaps, we conducted an authentic classroom experiment with secondary physics students in the context of inquiry learning to investigate whether the effectiveness of non-interactive teaching can be improved by 1) adding drawing as a visual-spatial generative activity to the act of verbal teaching (cf. drawing-facilitates-explaining hypothesis, Fiorella, 2023b), and by 2) distributing the generative activity across the study phase, as a concurrent activity (Bisra et al., 2018; Cuddy & Jacoby, 1982; Lachner et al., 2020). Additionally, we explored whether inter-individual differences might be associated with the effectiveness of non-interactive teaching.

4.2 Learning by Non-Interactive Teaching

An effective way to enhance student learning is to have them teach the previously learned contents to others, a method known as learning by teaching (Fiorella & Mayer, 2016;

Pi et al., 2021). Recently, researchers have begun to explore how *learning by teaching* can be effective even in non-interactive contexts. This so-called *non-interactive teaching* is a generative learning activity in which students are asked to generate an explanation to a fictitious peer of the previously learned contents (Lachner et al., 2022). Grounded in Wittrock's generative model of learning (1989, 2010), and related models such as the select-organize-integrate (SOI) model of generative learning (Fiorella & Mayer, 2016), non-interactive teaching aims at fostering students' meaningful learning by triggering active knowledge construction processes (Brod, 2021; Fiorella, 2023b; Fiorella & Mayer, 2016). During teaching, students engage in generative processes which are regarded as enhancing learning and metacomprehension. First, students need to select the most relevant information of the learning contents. Second, they need to organize the information into a coherent order to be able to generate a comprehensible explanation. Third, the students need to integrate the new information with their prior knowledge to be able to provide further details and examples that go beyond the given materials to address their explanation to the audience's needs. Throughout this process, non-interactive teaching may trigger metacomprehension to monitor one's own understanding. Importantly, even the mere expectancy to teach—without actually teaching—can contribute to students' learning by encouraging deeper processing and knowledge organization (Fiorella & Mayer, 2013, 2014; Guerrero & Wiley, 2021; Hoogerheide et al., 2014).

Several studies demonstrated beneficial effects of non-interactive teaching regarding students' learning and metacomprehension (Hoogerheide, Renkl, et al., 2019; Jacob et al., 2020; Lachner et al., 2021; Pi et al., 2021). Recent meta-analytical evidence showed small positive effects of non-interactive teaching regarding students' cognitive learning (Kobayashi, 2024: $g = 0.27$, small effect; Lachner et al., 2021: $g = 0.22$ for conceptual knowledge, $g = 0.16$ for transfer, both small effects; Ribosa & Duran, 2022: $g = 0.17$, small effect). Similarly, researchers found positive effects on metacomprehension (Fukaya, 2013; Jacob et al., 2020;

Lachner et al., 2020). Fukaya (2013), for instance, conducted an experiment in which university students were asked to read five texts with the intention to teach the contents afterwards. Then, students either taught the contents (teaching condition) or only planned to teach but did not generate an explanation (intention-only condition). The control group had the intention of writing keywords after reading the text and then actually wrote keywords. Results showed that students who taught demonstrated significantly better monitoring accuracy, meaning that they judged their knowledge in the test more accurately than those who only intended to teach or those who wrote keywords (main effect: $\eta^2 = 0.17$, large effect).

Notably, prior research predominantly took place in laboratory settings with university students with only immediate or short delay posttests (e.g., Fiorella & Mayer, 2013, 2014; Hoogerheide et al., 2014). Because laboratory settings might differ from authentic classrooms in terms of their study populations and contextual factors like more pronounced learner diversity, environmental distractions and social dynamics (e.g., Dinsmore & Alexander, 2012), it is unclear whether non-interactive teaching is also effective in authentic classrooms with school students (see also Sibley et al., 2024). Additionally, little is known about potential lasting learning effects of non-interactive teaching, as prior research predominantly focused on effects of teaching regarding students' immediate learning and not their lasting learning.

4.2.1 Enhancing the Effectiveness of Non-Interactive Teaching

Even though the potential benefits of non-interactive teaching have been documented by recent meta-analyses (Kobayashi, 2024; Lachner et al., 2021; Ribosa & Duran, 2022), their findings varied considerably among and even within studies. While some reported positive effects, others found null or even negative effects, indicating that non-interactive teaching is not necessarily effective. Recently, researchers have begun to explore two primary strategies for enhancing non-interactive teaching. First, they have combined non-interactive teaching with additional generative learning activities like drawing (e.g., Fiorella & Kuhlmann, 2020).

Second, they have investigated ways to improve the activity of non-interactive teaching (e.g., Lachner & Neuburg, 2019), such as optimizing the timing of the learning activities (i.e., distribution, see Lachner et al., 2020).

4.2.1.1 Enhancing Non-Interactive Teaching With Drawing

Drawing is regarded as a learning activity in which students are asked to draw the previously learned contents (Fiorella & Mayer, 2016). During drawing, students need to re-organize the information to be able to generate visual-spatial representations, which often surpassed the initially provided information and therefore can enhance students' learning and metacomprehension (Ainsworth & Scheiter, 2021; Fiorella, 2023; Van Meter & Firetto, 2013). Regarding students' learning, Fiorella's and Zhang's (2018) demonstrated in their meta-analysis that students benefited more from drawing than from reading or text-focused strategies such as summarizing or paraphrasing (d = 0.46 for comprehension, small effect; d = 0 .70 for transfer, medium effect). Regarding students' metacomprehension, Fiorella and Jaeger (2023), for instance, showed that students who either created their own visualizations or used those generated by instructors exhibited better monitoring accuracy than those who studied with only text or provided visuals (for explain judgments: d = 0.36, small effect).

According to the generative sense-making framework proposed by Fiorella (2023b), non-interactive teaching emphasizes generating coherent verbal explanations, which supports knowledge generalization. In contrast, drawing emphasizes creating an external visualization (Cox, 1999; Schmidgall et al., 2019; Van Meter & Firetto, 2013) that organizes information in a conceptually meaningful way. Non-interactive teaching and drawing place considerable demands on working memory when generating explanations and drawings, while drawing also enables students to offload and externalize their thoughts through visual representation, thereby possibly freeing up cognitive resources that students could invest in learning-relevant processes (Fiorella, 2023b; Fiorella & Mayer, 2016; Sweller et al., 2011). Fiorella (2023b) further suggested that combining these approaches can be particularly effective: the visualizing function of drawing can scaffold and strengthen the verbal explanation process, providing an external structure that aids in explanation and promotes deeper cognitive and metacognitive engagement.

To date, there is limited research regarding combining teaching and drawing. As an exception, Fiorella and Kuhlmann (2020) examined the influence of drawing and teaching with 120 college students who either taught previously learned contents, created drawings, taught and drew, or restudied the contents. Results of a one-week delayed posttest showed that drawing enhanced students' learning compared to restudying ($d = 1.06$, large effect), which was also true for teaching-only ($d = 0.80$, medium effect) and for combined teaching and drawing compared to restudying ($d = 1.46$, large effect). Moreover, students who taught and drew outperformed students in the drawing-only ($d = 0.65$, medium effect) and teaching-only condition ($d = 0.99$, large effect). Whether these findings can be transferred into authentic classrooms settings with school students is, however, still an open question.

4.2.1.2 Enhancing Non-Interactive Teaching Through Distribution

The timing of teaching may also be critical for the effectiveness of non-interactive teaching. Inspired by approaches of interpolated testing (Pan et al., 2024), distributing a teaching activity within a study session can be more effective than a single teaching activity at the end of a study phase, since multiple generative activity phases may involve multiple chances of generation processes that may contribute to the (re-)construction of knowledge (Cuddy & Jacoby, 1982; for meta-analytical evidence, see Prinz et al., 2020b; for related evidence, see Carpenter et al., 2012b; Ebersbach et al., 2022). From a metacognitive perspective, distributed learning activities may additionally aid monitoring the learning process and subsequent regulation activities by providing students with multiple opportunities to evaluate their progress, compare their understanding across phases, and adjust their strategies. For example, they can refine time management, prioritize contents, or address misunderstandings, which may improve their monitoring accuracy and learning regulation (Lachner et al., 2020; Schleinschok et al., 2017).

When distributed teaching is effective, it typically involves distributing explanations across multiple learning episodes, allowing students to revisit and reconstruct their knowledge over time. Cuddy and Jacoby (1982, Exp.1) demonstrated this effect in an experiment in which 18 university students learned word pairs, with the second word presented either intact or with missing letters (e.g., TREE: BR-CH), and recalled them after varying intervals. The authors manipulated the number of intervening tasks between repetitions, showing that recall performance improved when prior knowledge was less readily accessible due to a greater number of intervening tasks ($F(2,34) = 5.65$, $p < .050$). Specifically, recall was higher when four or eight unrelated items appeared between repetitions compared to immediate repetition, but this effect occurred only when students had to reconstruct the missing letters (main effect of repetition condition: $F(1,17) = 16.28$, $p < .050$; interaction distributing × repetition: $F(2,34) = 64.71$, $p < .050$). These findings suggest that distributing learning opportunities, combined with constructive learning processes, can facilitate deeper encoding and stronger memory retention.

However, most of the previous research on teaching and drawing, as well as educational practice, asked students to only teach or draw once, namely at the end of a study phase (Lachner et al., 2020). Thus, little is known about the effect of distributing non-interactive teaching. So far, Lachner et al. (2020) have been the only researchers to systematically distribute the teaching activity within their study. In two experiments, the authors investigated whether distributing non-interactive teaching at one point of time during students' studying (i.e., distributed teaching) would support learning more than non-interactive teaching after the entire study phase (i.e., no distributed teaching). The findings showed that students who taught once during studying achieved higher conceptual knowledge scores in an immediate knowledge test than those who taught once after the study phase (η_p^2 = .06, medium effect). The benefits of distributed teaching were explained by students' more frequent engagement in monitoring.

Although this study provides first evidence that distributing non-interactive teaching matters, it is still unclear whether distributing non-interactive teaching in several phases is even more beneficial, whether the effects of distributed teaching are transferable to classrooms, and whether they also may result in lasting learning effects, since the authors did not apply a delayed posttest.

4.2.2 Influence of Students' Individual Differences

Against the backdrop of the mixed findings of non-interactive teaching, in their theoretical review, Lachner et al. (2022) advocated for a thorough examination of boundary conditions of non-interactive teaching, particularly regarding students' prerequisites. Snow's Aptitude-Treatment-Interaction (ATI) theory (1991) provides a framework for understanding how individual differences can influence the effectiveness of instructional strategies. According to the ATI theory (Snow, 1991), the success of an instructional method, such as non-interactive teaching, depends on the alignment between the instructional approach and students' individual aptitudes. Specifically, students' prerequisites—such as prior knowledge, motivational or personality factors—can determine how effectively they engage in and benefit from generative learning activities. These individual differences may affect students' ability to select, organize and integrate new information with existing knowledge—generative processes required for effective learning (Fiorella, 2023b; Fiorella & Mayer, 2016).

For instance, prior knowledge is a crucial cognitive factor (Kalyuga, 2007; McNamara et al., 1996; J. Richter et al., 2018) that influences how students construct robust mental representations and efficiently integrate new information (Lachner et al., 2021; Mayer, 2009). The ability to form these connections is essential for meaningful learning (Fiorella, 2023b; Fiorella & Mayer, 2016). Students with limited prior knowledge may struggle to make these connections (Renkl, 2014), often resulting in suboptimal learning outcomes (Roelle & Nückles, 2019). In contrast, students with high prior knowledge may find generative learning strategies redundant, as they can manage learning more independently (Castro-Alonso et al., 2021). Thus, generative learning strategies, such as non-interactive teaching, may be particularly beneficial for students with lower prior knowledge, as they provide needed support (McNamara & Scott, 1999).

To date, only Hoogerheide, Renkl, et al. (2019) examined the influence of university students' prior knowledge regarding generative learning within non-interactive teaching settings. In their study, the researchers compared the performance of students engaged in teaching activities with those engaged in restudying, within the context of electrical troubleshooting tasks. Findings indicated that the teaching condition outperformed the restudy condition on both isomorphic (η_p^2 = 0.07, medium effect) and transfer problems (η_p^2 = 0.07, medium effect) in posttest performance. Regarding transfer, students with lower prior knowledge particularly benefited from teaching, while the teaching condition eliminated the positive relationship between prior knowledge and transfer performance observed in the restudy condition (β = 0.56, medium effect in the restudy control group).

Relatedly, prior research has highlighted academic self-concept as a fundamental prerequisite for learning. Academic self-concept describes how students perceive and evaluate their own competency within a particular academic field (Marsh et al., 2017; Shavelson et al., 1976). It is influenced by various factors, including experiences, feedback from teachers, and social comparisons. These comparisons are crucial in shaping students' beliefs about their academic competence, as they continuously assess their performance relative to that of their classmates or peers within the learning context (Marsh et al., 2008). Although academic self-concept has been shown to be positively correlated with learning outcomes (Möller et al., 2020; Valentine et al., 2004), its role in non-interactive teaching contexts remains underexplored. Given that academic self-concept influences how students approach and engage with learning tasks (Urhahne & Wijnia, 2023), understanding its interaction with instructional strategies like learning by teaching is critical. As an exception, Jacob et al. (2022) examined non-interactive teaching with seventh-grade school students who either taught the learning contents to a fictitious peer or were engaged in a retrieval practice. Results showed no effects among conditions. Interestingly, explorative analyses revealed that students with low academic self-concept benefited from teaching, possibly because they gained more from structured, generative activities like teaching. In contrast, students with high academic self-concept appeared to benefit more from retrieval practice, as they likely did not require additional scaffolding or structured learning activities to achieve optimal learning outcomes (interaction effect: $\beta = 0.49$, medium effect).

In addition to cognitive (i.e., prior knowledge) and motivational (i.e., academic self-concept) influencing factors, research also highlighted students' conscientiousness as a crucial individual pre-requisite for learning (Komarraju et al., 2009; Poropat, 2009; Richardson et al., 2012; Waldeyer et al., 2022; H. Wang et al., 2023). Conscientiousness includes self-control, responsibility, diligence, goal-directed behavior, effective planning, orderliness, delay gratification, and adherence to rules and norms (De Fruyt et al., 2008; John et al., 2008; John & Srivastava, 1999; McCrae & John, 1992; Roberts et al., 2014). There is ongoing debate about whether conscientiousness functions as a general trait or has domain-specific expressions (Mammadov, 2022; Meyer et al., 2023). Conscientiousness can also be operationalized in different ways, one of which is work ethic as a facet of this trait (Roberts et al., 2014; see also De Fruyt et al., 2008; Mang et al., 2018). Work ethic refers to the propensity for persistence, diligence, and hard work in goal-oriented tasks (Roberts et al., 2014). A strong work ethic could positively influence students' engagement in generative learning activities, as it may affect how thoroughly, precisely, and consistently students approach these tasks. According to the generative sense-making framework (Fiorella, 2023b), these behaviors can facilitate generative processes, for instance, students with higher levels of work ethic may be more likely to invest effort in thoroughly understanding new material and connecting it with prior knowledge. These qualities are particularly important in non-interactive teaching contexts, where students must independently regulate their learning behaviors. Therefore, a strong work ethic may enhance the effectiveness of generative learning strategies and contribute to improved learning outcomes and metacomprehension (e.g., Song et al., 2020 for related empirical evidence; Spielmann et al., 2022 for a comprehensive review). However, to our knowledge, no one has taken students' work ethic into account when analyzing non-interactive teaching.

4.3 The Present Study

In this study, we combined non-interactive teaching with drawing and varied its interpolation (after-study activity versus distributed throughout the study phase). Additionally, we explored students' individual differences (i.e., prior knowledge, academic self-concept, work ethic) as crucial moderating factors of non-interactive teaching. We tested the following preregistered hypotheses (https://aspredicted.org/RM1_1TF):

4.3.1 Generation Hypothesis (H1)

Based on the generative learning theory (Fiorella, 2023b; Fiorella & Mayer, 2016; Wittrock, 1989, 2010), we hypothesized that students who engage in non-interactive teaching (teaching-only, teaching + drawing) show better performance than students who restudy the learning contents (restudy control group) regarding their a) conceptual knowledge (Fiorella & Mayer, 2013, 2014; Hoogerheide, Visee, et al., 2019) and b) monitoring accuracy (Fukaya, 2013; Jacob et al., 2020). Additionally, we explored whether the effects also resulted in lasting learning (delayed test after eight weeks).

4.3.2 Drawing Hypothesis (H2)

Previous studies highlighted the benefit of learner-generated visualizations (Cooper et al., 2017) and combining teaching and drawing for knowledge enhancement (Fiorella, 2023; Fiorella & Kuhlmann, 2020). Accordingly, we hypothesized that students who teach and draw (teaching + drawing) outperform students who only teach the new contents (teaching-only) regarding their a) conceptual knowledge and b) monitoring accuracy. Additionally, we explored whether the effects also resulted in lasting learning (delayed test after eight weeks).

4.3.3 Distribution Hypothesis (H3)

Previous research showed that distributed learning activities can be more effective than learning activities after a study phase (Lachner et al., 2020). Therefore, we hypothesized that students in the distributed conditions would outperform students in the after-study conditions regarding their a) conceptual knowledge, and b) monitoring accuracy. Additionally, we explored whether the effects also resulted in lasting learning (delayed test after eight weeks). Furthermore, we explored potential interaction effects between the learning activity (i.e., restudy, teaching-only, teaching + drawing) and the timing factor (i.e., after-study, distributing).

4.3.4 Explorative Analyses on Students' Individual Differences

Following Lachner et al. (2022), we closely inspected students' individual differences, as they could critically influence their learning and metacomprehension (Hoogerheide, Renkl, et al., 2019; Jacob et al., 2022; Song et al., 2020). To replicate previous findings, we included students' prior knowledge (Hoogerheide, Renkl, et al., 2019) and academic self-concept (Jacob et al., 2022). As an extension, we included students' work ethic as a facet of conscientiousness, given its crucial role as a predictor of learning as evidenced by related research (Bareis et al., 2024; Song et al., 2020; Spielmann et al., 2022).

4.4 Method

4.4.1 Participants and Design

In total, 345 school students in the seventh and eighth grade from four secondary schools in south-west Germany participated in our study. As some school students ($n = 28$) only attended the pretest or the delayed test but were absent in the main part of the study (i.e., teaching unit), we excluded these data sets from the subsequent analyses, resulting in a total sample size of N

= 317 (for more details see Appendix A). This sample size exceeded the required sample size of 206 students, computed via an a-priori power analysis using G*Power (f = 0.25, α = 0.05, 1 – β = 0.90, ANCOVA with one covariate—prior conceptual knowledge or prior monitoring accuracy).[5]

The mean age of the final sample was 12.37 years (*SD* = 0.70) and 49.50% were female. Most of the students (60.47%) stated that German was their native language, 14.29% grew up bilingual with German, and 25.25% had another native language. The students showed relatively low prior cognitive knowledge (*M* = 7.80, *SD* = 3.6; max. 30 points), and moderately high levels of academic self-concept (*M* = 2.75, *SD* = 0.56) and work ethic (*M* = 2.96, *SD* = 0.50) in physics, as measured on Likert scales from one to four.

To test our hypotheses, we applied a 3 × 2 design with learning activity (restudy, teaching-only, teaching + drawing) and timing (after-study, distributed) as between-subject factors. Thus, students were either engaged in the learning activity only once, namely after the study phase (after-study: restudy: n = 53, teaching-only: n = 52, teaching + drawing: n = 53) or several times during the study phase (distributed: restudy: n = 48, teaching-only: n = 56, teaching + drawing: n = 55). The learning tasks were realized as open-book activities, allowing students access to the learning material. We implemented restudy as control condition, since restudy induces generative processes to a less pronounced extent (Fiorella & Mayer, 2016).

4.4.2 The Teaching Unit

The curriculum-aligned teaching unit in physics was about the converging lens and its images (geometrical optics), a standard topic for this age group. The teaching unit was taught by a certified physics teacher with 10 years of teaching experience (first author) who was also

---

[5] We based our power analysis on ANCOVA rather than planned contrasts to adopt a more conservative approach. As our analyses used multilevel modeling, power considerations should be interpreted with caution.

the researcher at the schools. All instructional resources used (i.e., introduction, overview sheet, experimentation worksheet) were based on Flegr et al. (2023).

4.4.2.1 Introduction to the Converging Lens

The introduction to the converging lens and its basic functions was delivered through a teacher-led presentation using visual slides. The presentation covered the use of lenses in daily life, prompting students to connect with their prior knowledge through questions such as "Where can you find lenses in your everyday life?", followed by a class discussion and examination of visual examples. Additionally, the introduction also provided a definition of converging lenses, key terms, and an explanation with an illustration of light refraction through a converging lens, along with basic details on the images formed by these lenses. An additional overview sheet outlined the fundamental concepts and functions of converging lenses (see first worksheet by Flegr et al., 2023).

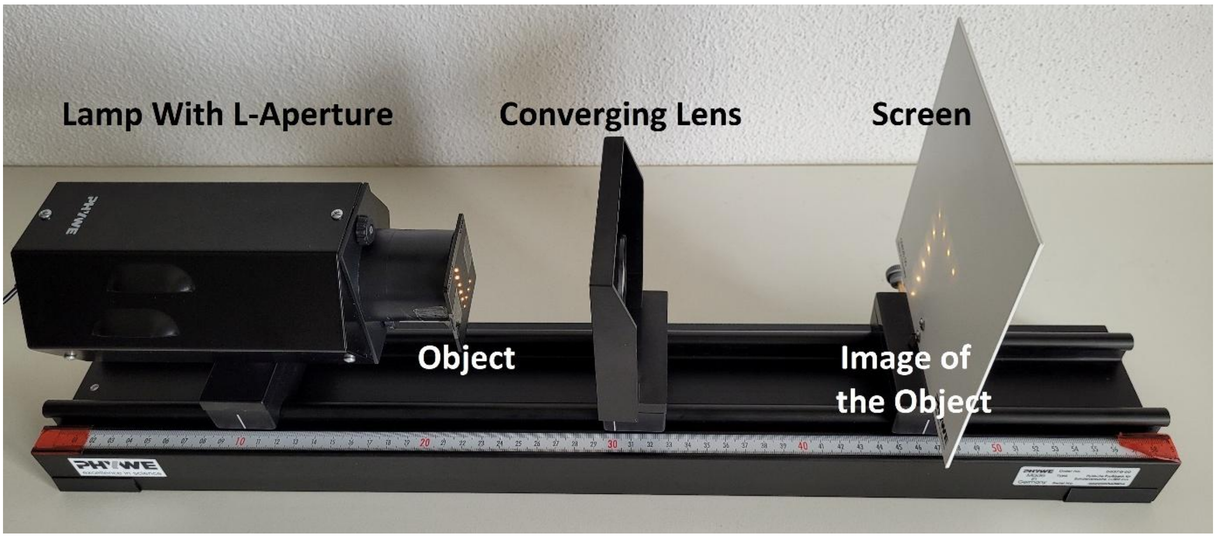
4.4.2.2 Students' Experiments

After the introduction, students examined how covering parts of a converging lens (affecting its diameter) and altering the distance between the object and the lens influenced the resulting image. During these experiments, students utilized an optical bench, an LED lamp with a "Perl-L" aperture, a converging lens, and a white screen. The light source gives a lighting "L"-shaped object, which could be moved closer to or further from the lens. The light from the "L"-shaped object refracts through the lens and forms an inverted image on a movable screen located opposite the lens (Figure 11).

A worksheet prompted the students to formulate hypotheses (e.g., "What do you think happens to the image of the object when the lens is partially covered?"), answer multiple-choice questions (e.g., "Is the complete 'L' still imaged when the lens is half covered? Yes, there is no difference. / Yes, but the 'L' on the screen is not as bright as before. / No, the 'L' is cut off. / No, the 'L' is no longer visible"), and answer an open-ended question ("Can you explain why that is?"). It also included a table for documenting observations (object distance, image distance, image characteristics), an exercise to select correct word elements (e.g., "If the lens is partly covered, an image of the complete object *is still formed / is not formed* on the screen"), and a section for comparing initial hypotheses with actual results (e.g., "Compare the mnemonic with your previously formulated hypothesis 1. Was your assumption different than the result? Yes, completely different. / A bit. / No, it was the same").

Figure 11

Students' Experimental Setup on the Converging Lens and its Images



4.4.3 Learning Activity

During the learning activity, students were asked to either teach the contents to a fictitious peer, to teach and draw a picture, or to restudy the contents. We focused on two main learning objectives of the study (covering parts of a converging lens and altering the distance between the object and the lens) which we addressed in the instructions. Students in the teaching conditions (teaching-only, teaching + drawing) were given the following instruction:

In a message, the student ***Mia*** *wrote to you* ***two assumptions*** *about the converging lens and its images.* ***Reply to Mia*** *by creating a* ***clear and detailed voice message to Mia*** *so that she can understand the contents without any additional information. Respond to* ***Mia's assumptions****. You are allowed to use the materials of the topic, but it is very important that you* ***DO NOT*** *read from them when recording the voice message but* ***formulate it in your own words*** *and incorporate your* ***own thoughts***[6]*. You have a total of* ***15 min*** *to complete this task. Be sure to use all the time.*

Students in the teaching-only condition then saw a mock-up chat on a tablet device with the fictitious peer Mia (see Figure 12, left side) in which Mia stated two assumptions ("I actually believe: When the lens is covered half, only half of the L is shown as an image."; "I actually believe: When I move the object towards the lens, I also have to move the screen closer to the lens in order to see a sharp image on the screen"), addressing common misconceptions about image formation with converging lenses (see Wörner et al., 2022). Students could teach Mia the contents by sending her a voice message.

Students in the teaching + drawing condition received the same text from Mia on a tablet device and could also teach by sending a voice message. In addition, they were provided with two template pictures in which they could draw during teaching (see Figure 12, right side).

To guarantee that students in the control group were provided with the same contents, we presented them with the following detailed instruction, in which also both main misconceptions were addressed:

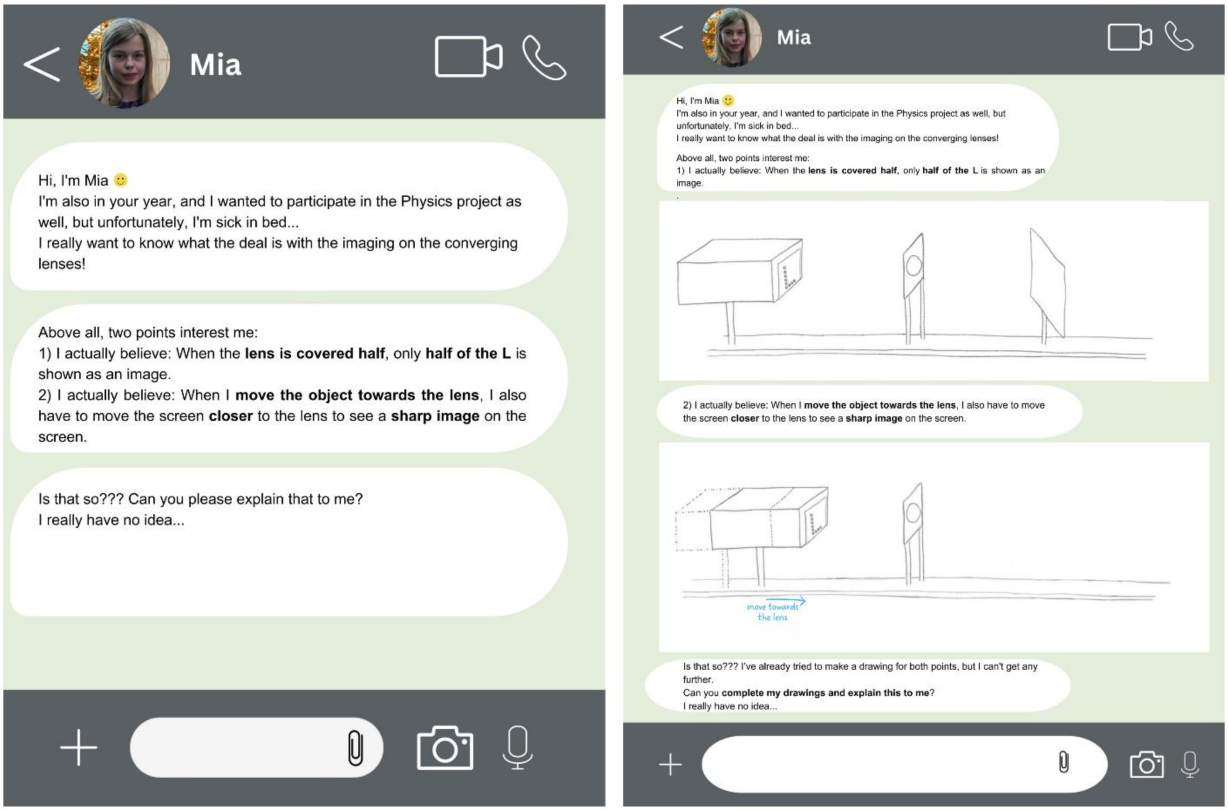
Restudy *the contents of today's physics lessons. To do this, use* ***today's physics materials****: You have received an overview sheet on the basics about the converging lens and its images and you have the instructions and results of the experiments. Use the materials for*

[6] Note that we checked students' adherence to these instructions and found no indications that students read directly from the materials instead of using their own words. Regarding students' drawings, no visualizations were provided that they could directly replicate.

restudying. ***Focus on the following two points:*** *What happens to the image if the lens is partially covered? What happens to the image if the distance between the object and the lens is changed? Take* ***notes*** *on this sheet. You may use the front and back of the sheet for this. You have a total of* ***15 min*** *to complete this task. Be sure to use all the time.*

Figure 12

Mock-up Messenger Chat for the Generative Conditions (Teaching-Only, Teaching + Drawing)



Note. The mock-up chat shows a message from the fictitious peer Mia, including a profile picture. Students in the teaching-only (left) and the teaching + drawing condition (right) could record a voice message. Translated from German.

4.4.4 Measures

Data collection was paper based. As this study was integrated within a larger research initiative, we only present the relevant variables for this study (see Appendix B, for detailed inventory of variables).

4.4.4.1 Conceptual Knowledge

To evaluate students' conceptual knowledge in the pre-post-delayed test, we utilized the validated conceptual knowledge test ROC-CI by Wörner et al. (2022). The test consisted of 15 multiple-choice items, designed to test students' conceptual knowledge about the operation of a converging lens, including distractors which reflected common misconceptions (e.g., "The arrow travels as a whole to the lens, is flipped by the lens, and then travels to the screen"). For a correct answer, students were awarded two points, for a partly correct answer with one point, resulting in a maximum score of 30 points. To minimize the possibility of recognition bias, we displayed answer choices in a random sequence for the pre-, post-, and delayed test. Two independent raters coded 20% of students' responses. The interrater reliability was excellent ($ICC_{2,1} = 1.00$), therefore only one of the raters coded the remaining answers. The reliability of the test was sufficient (McDonald's $\omega_t = 0.68$).

4.4.4.2 Monitoring Accuracy

To evaluate students' monitoring accuracy, we asked them to predict their expected performance on the pretest, the immediate posttest, and the delayed posttest. Each of these three predictions was made immediately before the corresponding test: "In the following you will answer 15 questions about the topic 'Imaging by converging lenses'. You can get two points per question. In total, you can score 30 points. How many points do you think you will get?" (Baars et al., 2017; Jacob et al., 2022; Prinz et al., 2018) on a scale from zero to thirty points (for a similar approach, see Jacob et al., 2020, 2022; Schleinschok et al., 2017). We quantified monitoring accuracy for the pretest, immediate posttest, and delayed posttest through a bias metric[7] (Lachner et al., 2020; Schraw, 2009), which measures the difference between the

[7] In our preregistration, we mistakenly used the word 'absolute' when providing a detailed explanation of the bias measure for monitoring accuracy. We have disclosed this wording error to ensure transparency and scientific rigor. This approach aligns with the recommendations by Lakens (2024), who emphasizes that deviations from a preregistration due to errors can be transparently addressed to improve the validity of the analysis.

respective predicted and actual scores (i.e., $X_{Judgment} - X_{Performance}$). Negative values (minimum: –30) indicate students' underestimation, positive values (maximum: 30) indicate their overestimation, and a zero-value represents an exact accuracy. We used bias instead of, for example, absolute accuracy (which measures the precision of a single judgment by calculating the squared deviation from actual performance; Schraw, 2009) because bias is a more informative measure of the direction and magnitude of metacognitive judgment errors (for further methods to measure monitoring accuracy, see e.g., Fiorella et al., 2024; Fiorella & Jaeger, 2023; Fukaya, 2013; Schraw, 2009). In our case, bias provides additional explorative critical insights into over- or under-confidence, which could be essential for understanding metacognitive processes (Schraw, 2009).

4.4.4.3 Academic Self-Concept in Physics

We determined students' academic self-concept in physics with four items (e.g., "I even understand the most difficult tasks in physics lessons") on a four-point Likert scale from one "I completely disagree" to four "I completely agree" (see Flegr et al., 2023; Mang et al., 2018). Reliability was good (McDonald's $\omega_t = 0.81$).

4.4.4.4 Physics Work Ethic

Physics work ethic was operationalized as a facet of the personality trait conscientiousness (see De Fruyt et al., 2008; Mang et al., 2018; Roberts et al., 2014). We measured students' physics work ethic with four items (e.g., "I am paying attention in physics lessons") on a four-point Likert scale from one "I completely disagree" to four "I completely agree" (see Mang et al., 2018). Reliability was acceptable (McDonald's $\omega_t = 0.73$).

4.4.5 Procedure

The Ethics Committee of the University of Tübingen and the Ministry of Education and Cultural Affaires of the State of Baden-Württemberg approved this study. Participation was voluntary and data collection was limited to those students who had obtained written consent by their legal guardians. The procedure is summarized in Table 4.

Table 4

Procedure of the Study

After-Study Conditions	Distributed Conditions
Pretest (approx. 1 week before the Teaching Unit)	
The Teaching Unit: Introduction (15 min)	
The Teaching Unit: Students' Experiments (30 min)	**Learning Activity** (5 min)
	The Teaching Unit: Students' Experiments (5 min)
	Learning Activity (5 min)
	The Teaching Unit: Students' Experiments (25 min)
Learning Activity (15 min)	**Learning Activity** (5 min)
Immediate Posttest	
Delayed Posttest (after approx. 8 weeks)	

Note. Bold items varied across experimental conditions (i.e., restudy, teaching-only, teaching + drawing).

Approximately one week before the teaching unit, the first author visited the schools to introduce the classes to the upcoming study. This visit also included administering the pretest (approx. 30 min) to evaluate the students' demographic information, prerequisites (e.g., academic self-concept in physics, physics work ethic), and their prior conceptual knowledge and monitoring accuracy.

One week later, the first author, who is also a certified and experienced physics teacher (10 years), conducted the main part of the study (henceforth mentioned as "teacher"). First, the students were introduced to the topic of the converging lens through a teacher-led presentation using visual slides and received the overview sheet (15 min). Afterwards, the students were randomly assigned to one of six conditions (after-study: restudy, teaching-only, teaching + drawing; distributed: restudy, teaching-only, teaching + drawing) within each class. Students in the after-study versus distributed condition were separated to the left and right side of the classroom with a classroom divider between.

After randomization, students in the after-study condition directly conducted the experiments (30 min) in small groups of usually four students. During this phase, communication between students and the teacher mirrored typical physics lessons, allowing students to seek help if they were struggling with a task. After reviewing the experimental results with their teacher to confirm their accurate documentation, these students either taught the contents to a fictitious peer or taught and drew a picture individually (15 min). Students in the after-study control condition restudied the contents in the same amount of time on their own (15 min). During this activity, each student was assigned a specific seat and equipped with sound-proof ear protectors to prevent disturbances from fellow students. All learning materials were available to the students during the task. This open-book format was employed to circumvent any "hidden" retrieval-practice effects, and to keep the retrieval processes constant across conditions (Roelle, Endres, et al., 2023; Sibley et al., 2022). Moreover, it has been demonstrated that open-book teaching can be more effective than closed-book teaching (Sibley et al., 2022).

After the randomization described above, students in the distributed conditions directly either taught the contents to a fictitious peer, taught and drew a picture, or restudied the contents individually (5 min). Subsequently, they started to conduct the first part of the experiments (hypotheses generation, 5 min) also in small groups of usually four students and then again either taught, taught and drew, or restudied the contents individually (5 min). Next, these students engaged in the next experimental phase with the same small group (25 min) and reviewed the experimental results with their teacher. Again, the students in the distributed conditions either taught, taught and drew, or restudied the contents on their own (5 min), which kept the total time on task constant across all conditions (in total 15 min).

Afterwards, all students answered the immediate posttest (approx. 20 min). Importantly, students' regular physics teachers were instructed not to further address the concept of the converging lens and its images during the following eight weeks. At the end of this period, the delayed posttest was administered by the first author.

4.4.6 Analysis of Students' Experimentation Worksheets

To ensure high implementation fidelity, we assessed whether all students fully and correctly completed their experimentation worksheet, which was adapted from Flegr et al. (2023). Full and correct completion was coded as 1 (yes), while incomplete or incorrect completion was coded as 0 (no). This evaluation ensured that all students had the same basis for the subsequent learning activity.

4.4.7 Analysis of Students' Explanations, Drawings, and Restudy Notes

Students in the teaching conditions (i.e., teaching-only, teaching + drawing) taught the contents to the fictitious peer "Mia", and their explanations were analyzed for completeness, elaboration, and correctness to capture the underlying cognitive processes (for a similar approach, see Fiorella & Kuhlmann, 2020; Hoogerheide, Renkl, et al., 2019; Jacob et al., 2020, 2022). Restudy notes, along with students' drawings in the teaching + drawing condition, were also evaluated using these same indicators (see also Ainsworth & Scheiter, 2021; Fiorella & Kuhlmann, 2020; Schmidgall et al., 2019; Schwamborn et al., 2010). For students in the distributed conditions, completeness, elaboration, and correctness were assessed based on the aggregated final product across all learning activity segments to provide a comprehensive analysis (for a similar approach, see Fiorella, 2022). The detailed analysis scheme for students' explanations, drawings, and restudy notes is available here:
https://doi.org/10.17605/OSF.IO/UK9CH.

4.4.7.1 Completeness of Students' Explanations, Drawings, and Restudy Notes

The completeness of students' outputs was coded by identifying the concepts necessary to convey the learning contents: in the explanations and restudy notes, the concepts mentioned were coded; in the drawings, the concepts visually represented. Students could receive up to five points for each of the two task components (see Figure 12), resulting in a maximum of ten points in total (explanations and restudy notes: e.g., one point for "The image is swapped left-right and top-bottom compared to the object"; drawings: e.g., one point for depicting a complete image on the screen despite a partially covered lens; self-developed coding scheme based on conceptual knowledge of the converging lens, see Wörner et al., 2022).

Independent raters coded 20% of the explanations, drawings, and restudy notes. Interrater reliabilities were excellent (first part of the task: $ICC_{2,1} = 0.98 - 0.99$, second part: $ICC_{2,1} = 0.99 - 1.00$), so the remaining outputs (i.e., explanations, drawings, restudy notes) were split equally among the raters. The completeness scores from the two parts of the task were each added together, and this total value for completeness was used for further analysis of each teaching, drawing, or restudy note.

4.4.7.2 Elaboration of Students' Explanations, Drawings, and Restudy Notes

We counted the number of elaborations in each explanation, drawing, and restudy note to assess the level of elaboration. This included idea units such as analogies, examples, and personal experiences (see also Fiorella & Kuhlmann, 2020; Jacob et al., 2020; Lachner et al., 2018). For instance, the statement, "If I cover the lens ring-shaped, the image on the screen is also simply less bright than before," qualifies as an elaboration because it introduces information not presented during the learning phase. Similarly, a drawn light ray tracing the path from a specific object point of the "L" through the converging lens to its corresponding image point is considered an elaboration, as this detail was also not provided during the learning phase.

Independent raters assessed the number of elaborations for 20% of the explanations, drawings, and restudy notes. Interrater reliabilities were excellent for both parts of each task (first part: $ICC_{2,1} = 0.98 – 1.00$, second part: $ICC_{2,1} = 0 .96 – 1.00$). Thus, the remaining outputs were evenly split among the raters for coding. A total score for the number of elaborations per explanation, drawing, or restudy note was then calculated and used for further analysis.

4.4.7.3 Correctness of Students' Explanations, Drawings, and Restudy Notes

As an indicator of the level of *correctness*, we evaluated the percentage of correct knowledge reflected in the explanations, drawings, and restudy notes (Hoogerheide, Renkl, et al., 2019). First, we counted the idea units in both parts of each task. For example, in an explanation or restudy note, an idea unit was the statement "The image is less bright than before"; in a drawing, a drawn upside-down image of the "L". Next, we coded the number of physically correct idea units. Finally, we calculated the percentage of correct idea units for each part of the respective task.

Independent raters assessed the percentage of correct idea units for 20% of students' outputs. Interrater reliabilities were excellent for both parts of the respective task (first part: $ICC_{2,1} = 0.94 – 1.00$, second part: $ICC_{2,1} = 0.96 – 1.00$). The remaining outputs were then evenly divided among the raters for coding. A total score for the percentage of correct idea units was calculated for each explanation, drawing, and restudy note and used for further analysis.

4.4.8 Data Analyses

To test our preregistered hypotheses, we used planned contrasts while controlling students' prior conceptual knowledge or prior monitoring accuracy (Table 2). Conducting our preregistered contrast analyses instead of ANCOVAs allowed us to precisely test our hypotheses and their specific patterns (e.g., cascaded trends) while reducing the risk of alpha-inflation, as only one test was required (Furr & Rosenthal, 2003; Rosenthal & Rosnow, 1985). Moreover, contrast analyses are particularly recommended when testing specific hypotheses that cannot be adequately addressed by conventional ANOVAs (e.g., main effects and interactions), such as cascaded or synergistic trends (Wiens & Nilsson, 2017) and also require smaller sample sizes compared to conventional ANOVAs. To address the hierarchical structure of the data, with students nested within classes and classes nested within schools, we applied multilevel modeling[8] and incorporated cluster-robust standard errors. In the first contrast, we tested whether generation was more effective than restudy (i.e., restudy: –2; teaching-only: –1; teaching + drawing: 1). In the second contrast, we examined whether combining non-interactive teaching with drawing (i.e., teaching + drawing) was more beneficial than teaching-only (i.e., restudy: 0; teaching-only: –1; teaching + drawing: 1). Finally, in the third contrast, we tested whether the distributed learning activities were more effective than the after-study learning activities (i.e., after-study: –1; distributed: 1).

We explored potential interactions between the contrasted learning activities (i.e., restudy, teaching-only, teaching + drawing) and the timing factor (i.e., after-study, distributed) using the same planned contrast analyses conducted for the main effects. This approach allowed us to examine both main effects and interactions without the risk of alpha inflation associated with running multiple tests (Furr & Rosenthal, 2003; Rosenthal & Rosnow, 1985; Wiens & Nilsson, 2017). We also performed mediation analyses to explore potential mediation effects of the characteristics of students' explanations and restudy notes, using the contrast-coded experimental conditions as the independent variable. Additionally, we computed moderation analyses with the contrast coded predictors to test potential moderation effects of prior knowledge, academic self-concept, and work ethic.

[8] Including an additional level for student group did not change the results; the outcomes remained stable. Therefore, we chose a multilevel model with the levels class and school to provide a more ecologically valid analysis (see Greenland, 2000; Raudenbush & Bryk, 2002).

As our field study comprised several measurement time points, missing values naturally occurred. In the preliminary and main analyses, the missing values were managed by applying multiple imputations across 50 datasets and 50 iterations.

Data were analyzed using R, version 4.4.1 (R Core Team, 2024).

4.5 Results

The preregistration, data, analyses, and supplementary material are available here: https://doi.org/10.17605/OSF.IO/UK9CH. Cohen's d, partial η_p^2, and φ were employed to measure effect sizes (small effects: $d = 0.20$, $\eta_p^2 = 0.01$, $\varphi = 0.10$; medium effects: $d = 0.50$, $\eta_p^2 = 0.06$, $\varphi = 0.30$; large effects: $d = 0.80$, $\eta_p^2 = 0.14$, $\varphi = 0.50$, see Cohen, 2013). The alpha level was set to $\alpha = 0.05$.

4.5.1 Preliminary Analysis

Our examination for potential outliers through boxplot visualization did not reveal any outliers. Further preliminary analyses showed no significant prior differences between the conditions concerning gender, χ^2 (10, 317) = 6.83, $p = .732$, $\varphi = 0.14$, and first language, χ^2 (10, 317) = 8.71, $p = .563$, $\varphi = 0.17$. Separate one-way ANOVAs across the six conditions (restudy after-study, restudy distributed, teaching-only after-study, teaching-only distributed, teaching + drawing after-study, teaching + drawing distributed) indicated that conditions did not differ in students' age, physics work ethic, academic self-concept in physics, prior conceptual knowledge, and prior monitoring accuracy, $.670 < p < .797$. Table 5 shows the mean scores and standard deviations per condition. Additionally, our implementation fidelity check confirmed that all 317 students fully and correctly completed the experimentation worksheet. As a further implementation check, comparisons between the characteristics of the restudy notes and the generated explanations, with regard to completeness, elaboration, and correctness, showed that the explanations were more complete ($\beta = 0.20$, $p < .001$, small effect), contained more elaborations ($\beta = 0.11$, $p < .001$, small effect), but had no higher proportion of correct idea units ($\beta = -0.05$, $p = .195$) than the restudy notes. The addition of drawing to the teaching component did not result in significant differences regarding completeness ($\beta = -0.04$, $p = .558$) and elaboration ($\beta = 0.08$, $p = .307$), but did in correctness ($\beta = 0.22$, $p = .001$, small effect). Similarly, the timing did not affect the level of completeness ($\beta = 0.00$, $p = .953$) and elaboration ($\beta = 0.08$, $p = .148$), but did affect correctness ($\beta = -0.22$, $p < .001$, small effect), see also Table 5.

Table 5

Means and Standard Deviations for all Measurements Across Experimental Conditions

Variable	After-Study						Distributed					
	Restudy		Teaching-Only		Teaching + Drawing		Restudy		Teaching-Only		Teaching + Drawing	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Students' prerequisites												
Physics work ethic (1-4)	2.93	0.49	2.98	0.44	3.02	0.57	3.02	0.56	2.87	0.53	2.93	0.44
Academic self-concept in physics (1-4)	2.73	0.53	2.83	0.48	2.84	0.59	2.71	0.54	2.60	0.65	2.77	0.58
Prior												
Conceptual knowledge (0-30)	7.28	3.12	7.58	3.64	7.35	3.35	7.85	3.28	8.55	4.19	7.75	3.27
Monitoring accuracy[a] (-30-30)	9.94	6.95	9.94	6.37	9.84	6.94	9.93	6.88	8.06	7.31	8.83	5.30
Characteristics of students' explanations and restudy notes												
Completeness (0-10 points)	3.10	2.37	4.86	2.71	5.08	2.94	3.50	2.07	5.10	2.42	4.40	2.76
Elaboration (each 1 point)	2.12	2.50	3.42	3.33	3.60	7.53	2.67	2.12	3.70	3.67	4.98	6.02
Correctness (percentage)	94.84	9.78	91.98	13.13	94.04	9.16	89.58	14.15	79.58	19.88	90.61	10.77
Characteristics of students' drawings												
Completeness (0-10 points)	–	–	–	–	6.32	2.42	–	–	–	–	5.49	2.89
Elaboration (each 1 point)	–	–	–	–	1.96	2.77	–	–	–	–	2.65	3.35
Correctness (percentage)	–	–	–	–	87.13	19.35	–	–	–	–	77.37	23.74
Immediate												
Conceptual knowledge (0-30)	13.94	4.92	16.54	5.65	17.77	5.06	15.42	5.11	15.48	4.95	16.04	5.43
Monitoring accuracy[a] (-30-30)	2.08	7.94	-0.92	7.73	-1.92	6.94	0.44	6.83	-0.54	7.21	-1.09	7.66
Lasting												
Conceptual knowledge (0-30)	10.74	3.87	12.21	4.36	13.20	4.59	11.61	4.65	11.59	4.95	12.47	5.48
Monitoring accuracy[a] (-30-30)	2.69	6.16	0.00	6.92	0.14	7.07	2.50	6.89	1.69	7.73	1.04	6.61

Note. The data in this table are based on the raw data.

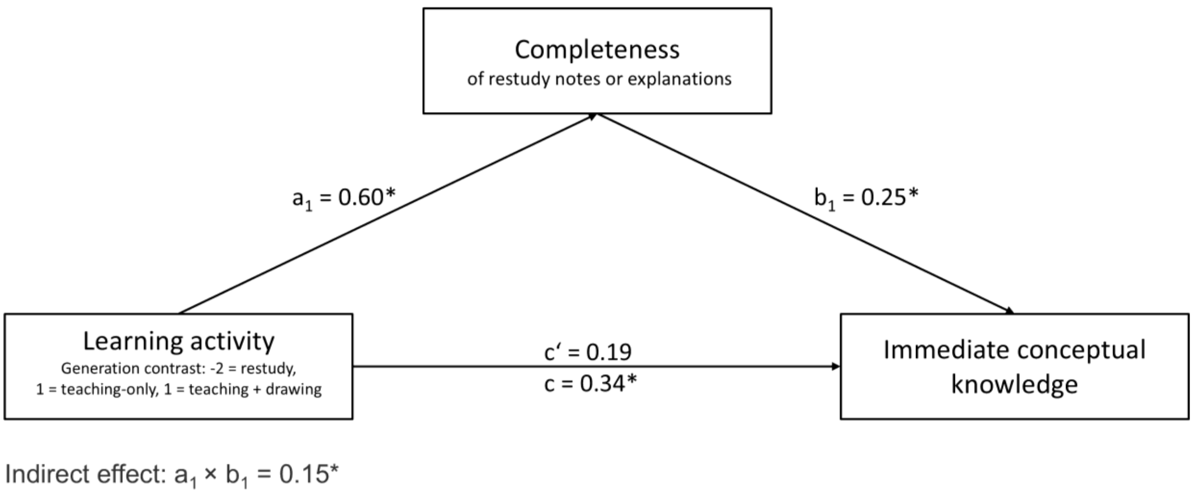
[a]Monitoring accuracy is calculated based on the differences between students' estimated and actual performance. Negative values indicate underestimation, positive values overestimation, and a value of zero a perfect judgment of learning.

4.5.2 Conceptual Knowledge

In line with our generation hypothesis (H1a), contrast analysis concerning the immediate posttest showed that students who engaged in non-interactive teaching (i.e., teaching-only, teaching + drawing) had higher conceptual knowledge than students who restudied the contents ($\beta = 0.11$, $p = .005$, small effect; see Appendix C for detailed results). To explore whether cognitive mechanisms underlie this significant generation effect contrasted to restudy, we performed explorative separate mediation analyses. The learning activity served as the predictor (generation contrast: –2 = restudy, 1 = teaching-only, 1 = teaching + drawing), with completeness, elaboration, and correctness acting as individual mediators, and students' immediate conceptual knowledge as dependant variable. The analysis revealed a significant indirect effect through completeness, $a_1 \times b_1 = 0.15$, $p = .001$ (see Figure 13 for the full mediation model). Neither elaboration nor correctness were significant mediators (for detailed results, see Appendix D).

Figure 13

Explorative Mediation Analysis in Terms of the Level of Completeness



Note. *$p < .050$. Regression coefficients are standardized.

The further contrasts were not significant, neither for immediate nor for lasting conceptual knowledge (see Appendix C for detailed results). This finding indicates that non-interactive teaching contributed to immediate learning but not to lasting learning. Additionally, neither drawing nor distributing generative activities contributed to students' conceptual knowledge.

In a next step, we also explored potential interactions between the contrasted learning activities (i.e., restudy, teaching-only, teaching + drawing) and the timing factor (i.e., after-study, distributed) using the same planned contrast analyses conducted for the main effects. In terms of immediate conceptual knowledge, we found a significant interaction between the contrasted restudy and generative conditions and timing ($\beta = -0.09$, $p = .017$, small effect; see Appendix C for further details). To break up this significant interaction, we conducted simple effect analyses regarding immediate conceptual knowledge. Results revealed that students who taught the learning contents only once (after the study phase) significantly outperformed students who taught distributed ($\beta = -0.30$, $p = .026$, small effect; after-study teaching: $M = 17.15$, $SD = 5.37$; distributed teaching: $M = 15.76$, $SD = 5.18$). In contrast, there was no effect of timing when students restudied the learning contents ($\beta = 0.25$, $p = .175$; after-study restudying: $M = 13.94$, $SD = 4.92$; distributed restudying: $M = 15.42$, $SD = 5.11$). Together, the results suggest that teaching had an effect on immediate conceptual knowledge only when teaching was conducted after-study.

4.5.3 Monitoring Accuracy

Regarding immediate monitoring accuracy, there was a significant difference between the restudy condition and the generative conditions (i.e., teaching-only, teaching + drawing), $\beta = -0.09$, $p = .013$ (small effect). Based on the descriptive values, students in the restudy condition overestimated their knowledge ($M = 1.29$, $SD = 7.44$), however, a one-sample t-test showed that their judgments did not significantly differ from zero, $t(100) = 1.72$, $p = .090$. In contrast, students in the generative conditions underestimated their knowledge (M = –1.11, SD = 7.36), with a one-sample t-test indicating a significant difference from zero, $t(215) = -2.21$, $p = .029$.

None of the other contrasts were significant, neither for immediate nor for lasting monitoring accuracy (see Appendix C for detailed results).

4.5.4 Explorative Analyses on Students' Individual Differences

Given that our intervention demonstrated effects predominantly regarding immediate conceptual knowledge, we explored students' prior knowledge, academic self-concept, and work ethic as moderators of our learning activities in the immediate posttest. To ensure a comprehensive understanding, we also conducted the corresponding analyses regarding lasting conceptual knowledge. We utilized separate analyses, since running one analysis including correlated concepts could result in lower statistical power, unstable regression coefficients, and larger error terms (Aguinis, 1995). For detailed results see Appendix E (for correlations see Appendix F).

4.5.4.1 Prior Knowledge

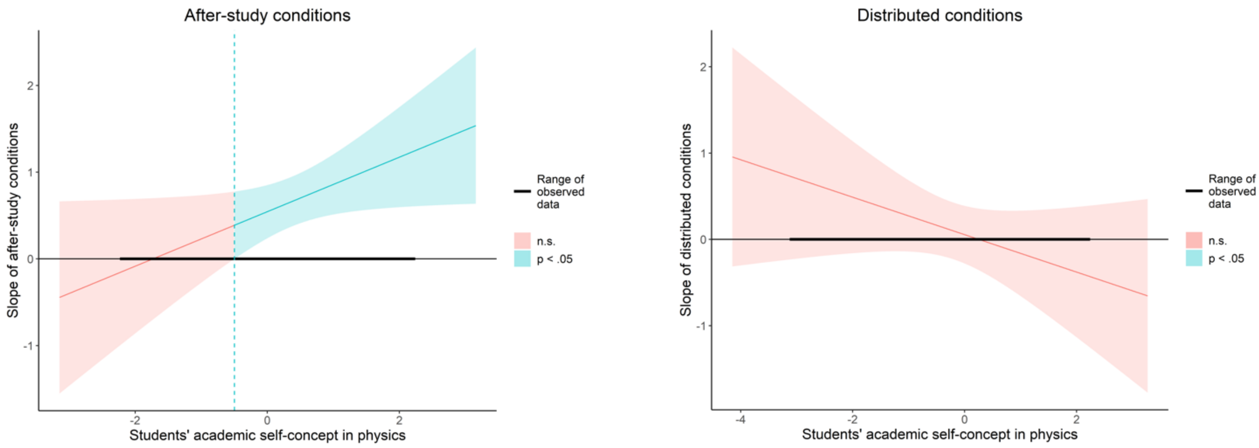
For immediate conceptual knowledge, we did not find a statistically significant interaction between students' prior knowledge and generation ($\beta = -0.04$, $p = .404$), nor between prior knowledge and drawing ($\beta = 0.02$, $p = .742$), nor between prior knowledge and distribution ($\beta = 0.07$, $p = .299$), nor between prior knowledge, generation, and distribution ($\beta = 0.04$, $p = .345$), and nor between prior knowledge, drawing, and distribution ($\beta = 0.04$, $p = .577$). None of the interactions for lasting conceptual knowledge were statistically significant.

4.5.4.2 Academic Self-Concept

For immediate conceptual knowledge, results revealed no significant interaction effects between students' academic self-concept and generation ($\beta = 0.02$, $p = .621$), not between academic self-concept and drawing (β = 0.03, p = .726), and not between academic self-concept and distribution (β = –0.01, p = .859). Interestingly, there was a statistically significant interaction between academic self-concept, generation, and distribution (β = –0.08, p = .026, small effect). To examine this interaction effect in more detail, we applied the Johnson-Neyman technique (Hayes & Montoya, 2017) which allows us to detect the significant range of students' academic self-concept in which the interaction is still significant. For the after-study condition, results revealed that students significantly benefited more from the generative activities than from restudying, however, only when their academic self-concept was medium to high (> –0.50). For students with a low academic self-concept (≤ –0.50), there was no difference among conditions. For the distributed condition, students' academic self-concept did not moderate the learning effect. Thus, academic self-concept was only a determining factor when students only taught once after the study phase.

Figure 14

Johnson-Neyman Confidence Intervals for Academic Self-Concept Moderation of Immediate Conceptual Knowledge in After-Study and Distributed Conditions



Note. The band reflects the 95% confidence interval which can be used to demonstrate significant regions. Significant regions are highlighted in blue, non-significant regions in red.

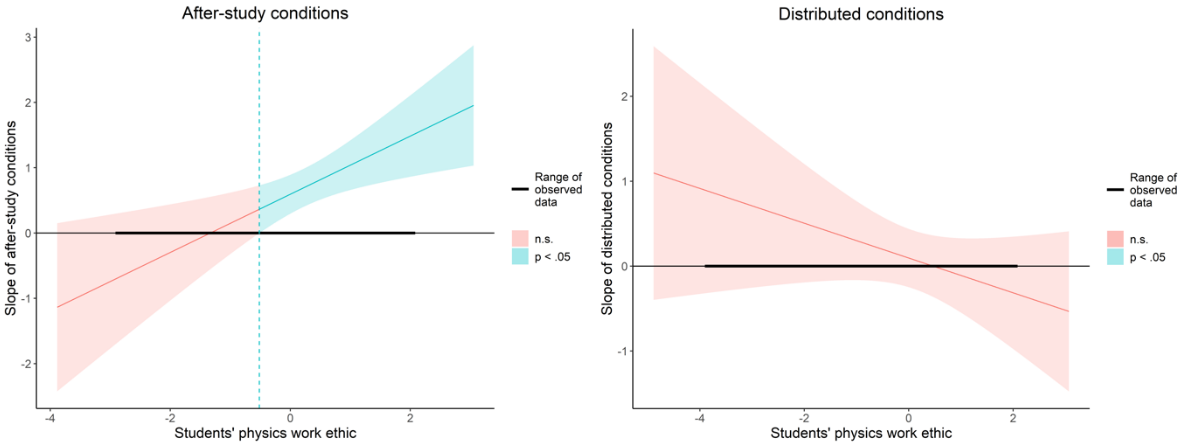
There was no statistically significant interaction between academic self-concept, drawing, and distribution (β = 0.02, p = .780). None of the interactions for lasting conceptual knowledge were statistically significant.

4.5.4.3 Physics work ethic

We did not find a statistically significant interaction effect of students' work ethic and generation (β = 0.03, p = .371), and not regarding work ethic and drawing (β = 0.13, p = .066) or work ethic and distribution (β = 0.00, p = .952). Again, there was a statistically significant interaction between work ethic, generation, and distribution (β = –0.10, p = .009, small effect). Regarding the after-study condition, the Johnson-Neyman technique revealed that for students with a physics work ethic of higher than –0.51, the generative activities were significantly better than restudying, whereas there was no significant difference among conditions for students with a work ethic of exactly and lower than –0.51. Regarding the distributed condition, there was no statistically significant interaction effect, indicating that students' work ethic only moderated after-study teaching.

Figure 15

Johnson-Neyman Confidence Intervals for Physics Work Ethic Moderation of Immediate Conceptual Knowledge in After-Study and Distributed Conditions



Note. The band reflects the 95% confidence interval which can be used to demonstrate significant regions. Significant regions are highlighted in blue, non-significant regions in red.

There was no statistically significant interaction between work ethic, drawing, and distribution ($\beta = -0.07$, $p = .359$). None of the interactions for lasting conceptual knowledge were statistically significant.

4.6 Discussion

In this field experiment, we investigated whether the effectiveness of non-interactive teaching can be improved by a) incorporating drawing and b) distributing non-interactive teaching across the study phase with regard to students' conceptual knowledge and monitoring accuracy. Additionally, we examined cognitive (i.e., prior knowledge), motivational (i.e., academic self-concept), and personality (i.e., work ethic) factors as moderators of non-interactive teaching.

For immediate conceptual knowledge, in line with our generation hypothesis (H1a), we found that students who engaged in non-interactive teaching outperformed those who restudied the contents, mediated by the level of completeness. Contrary to our drawing hypothesis (H2a), however, drawing did not significantly contribute to the effectiveness of non-interactive teaching. That said, we did not find a main effect of distributing the non-interactive teaching activities (H3a). However, explorative analyses of potential interactions between non-interactive teaching and distributing suggested that teaching had an effect only when conducted after-study but not when distributed throughout the teaching unit. Our explorative analyses on students' individual differences showed that academic self-concept in physics and physics work ethic were associated with students' immediate conceptual knowledge. In the after-study conditions, students with higher levels of academic self-concept or work ethic benefited more from the generative conditions compared to restudying. In contrast, no moderating effects were observed for students with lower levels of academic self-concept or work ethic. Thus, academic self-concept and work ethic were only associated factors when teaching occurred after the study phase. We did not observe significant lasting conceptual knowledge effects. For immediate monitoring accuracy, students in the restudy condition numerically overestimated their knowledge, while those in the generative conditions (teaching-only, teaching + drawing) significantly underestimated their performance. None of the other effects were significant, including those related to lasting monitoring accuracy.

From a theoretical perspective, our findings add to the scarce evidence that non-interactive teaching is a beneficial learning strategy in the context of inquiry learning in schools (e.g., Fiorella & Mayer, 2014; Hoogerheide, Renkl, et al., 2019; Kobayashi, 2024; Lachner et al., 2021; Palinscar & Brown, 1984). However, adding drawing to non-interactive teaching did not enhance students' understanding of the learning contents. At first glance, this result seems to contrast with the recent study by Fiorella and Kuhlmann (2020). However, Fiorella and Kuhlmann (2020) conducted their study in a laboratory setting with adults and informed them about the general task of the subsequent learning activity even before the study phase (i.e., teaching expectancy).

Further, distributing non-interactive teaching during the study phase was not more beneficial than teaching only once after studying. This finding contradicts our hypothesis and prior research which showed that distributing non-interactive teaching enhanced students' understanding (Lachner et al., 2020). However, in the study by Lachner et al., (2020), the authors implemented a distributed non-interactive teaching task only once, namely in the middle of the study phase. In contrast, we implemented three distributed teaching tasks during the study phase. Thus, our study extends prior research as our results indicate that enhancing non-interactive teaching through distribution at multiple points during the study phase did not optimize school students' learning. Nevertheless, it is unclear why teaching once (in the middle of the study phase) led to better outcomes but teaching several times during the study phase did not. One possible explanation is that frequent task-switching—between the respective learning activity and the ongoing teaching unit—may not always be beneficial. While one might assume that distributing learning activities benefits learning (Pan et al., 2024), it could also potentially increase the demands on students' working memory, particularly during knowledge construction. Future studies should investigate whether frequent interruptions in the learning process outweigh the benefits of distributed teaching and how to structure distribution in a way that optimally supports learning.

Interestingly, our analyses on students' individual differences revealed that prior knowledge did not moderate teaching and distributing as expected. Therefore, our findings are in line with other research that was unable to replicate such an interaction (Jacob et al., 2022; J. Richter et al., 2022). We attribute this finding to the fact that generative activities themselves are adaptable and can be worked on different levels of prior knowledge (Brod, 2024).

Additionally, our analyses revealed that academic self-concept moderated after-study learning activities regarding the acquisition of immediate conceptual knowledge. Students with high academic self-concept appeared to benefit more from generative activities than from restudying. A strong academic self-concept likely enabled these students to approach the complex and challenging physics topic (Flegr et al., 2023; Wörner et al., 2022) with confidence, possibly facilitating generative processes essential for effective learning (Fiorella, 2023b). In contrast, students with lower academic self-concept may have lacked the confidence or belief in their abilities to engage deeply with the task, reducing the effectiveness of teaching as a generative learning strategy. Notably, Jacob et al. (2022) found that students with low academic self-concept benefited more from teaching than retrieving. However, we want to note that Jacob et al. (2022) conducted their study in biology. Given that academic self-concept is domain-specific, our findings suggest that such ATI-effects may also depend on the particular domain in which non-interactive teaching is implemented (see also Sibley et al., 2024).

To our knowledge, we were the first to investigate the moderating role of physics work ethic regarding non-interactive teaching. Our findings demonstrated that work ethic, as a facet of conscientiousness, moderated after-study activities regarding immediate conceptual knowledge, which is also in line with previous publications in related areas (Bareis et al., 2024; Rieger et al., 2022; Song et al., 2020). We want to make explicit that we deliberately did not explore potential moderated mediations due to our restricted sample size and the fairly complex 3×2-design.

4.6.1 Study Limitations and Future Research

One limitation concerns the authenticity of our experimental field study. While our study was aligned with the physics curriculum and all lessons and study components were conducted personally by the first author (researcher and experienced physics teacher), the experimental setting may have dampened natural classroom dynamics, making it difficult to fully generalize the findings to real classroom teaching. Moreover, to isolate the lasting effects of the interventions, regular physics teachers were explicitly instructed not to revisit the topic of the converging lens and its images during the eight weeks between the immediate and delayed posttest. However, this differs from typical classroom practice, where topics are often periodically reinforced. Additionally, students may have independently engaged with the topic, introducing a potential confounding factor. Future research could enhance ecological validity by embedding interventions more seamlessly into standard classroom routines, thereby supporting the long-term integration of evidence-based teaching strategies. Additionally, we focused exclusively on physics. Given that prior research suggests non-interactive teaching is not equally effective across subjects or domains (Sibley et al., 2024) and that age-related differences exist (Brod, 2021), the generalizability of our findings remains uncertain.

Further limitations arise from the study design. First, while using the same conceptual knowledge test at three time points ensured consistency, it may have introduced recognition effects. To mitigate this, we randomized the order of answer options to prevent students from relying solely on memory. Nonetheless, future studies could explore alternative test formats to minimize potential biases. Second, we did not include a drawing-only condition, as our primary goal was to investigate whether adding drawing enhances the effectiveness of non-interactive teaching. Additionally, it would be valuable to replicate our findings under conditions where students are informed before studying that they will teach, as previous research suggests that this expectation can influence learning outcomes (Fiorella & Mayer, 2014; Guerrero & Wiley, 2021; Kobayashi, 2024).

It is also important to note that the observed effects were relatively modest, which may be attributed to the short duration of the intervention. Students engaged in generative tasks (i.e., teaching-only, teaching + drawing) only once for 15 min or in three separate five-minute segments. These brief intervals may not have provided sufficient time for students to fully engage with the teaching experience as a generative activity. Although each five-minute segment began only after all students were ready, we cannot rule out that some students may have needed additional time to process the material, potentially reducing overall time-on-task. Moreover, a total exposure of 15 min across all conditions may not have been sufficient to foster lasting learning effects over an eight-week period. Nonetheless, we want to note that our obtained findings are consistent with prior meta-analytic evidence (e.g., Lachner et al., 2022; Ribosa & Duran, 2022).

4.7 Conclusion

Our results demonstrated that non-interactive teaching was an effective generative strategy for supporting students' learning in school, likely because it led to a higher level of completeness compared to restudying. However, incorporating drawing or distributing non-interactive teaching at several time points during the learning phase did not yield additional learning benefits. Moreover, our study highlights the need to consider non-cognitive learner characteristics, as they can shape the non-interactive teaching effect. These insights provide valuable directions for optimizing non-interactive teaching and highlight the importance of considering individual differences in educational practice.

5

STUDY 3

HAPPY OR REDUNDANT? RETRIEVAL PRACTICE ENHANCES LASTING LEARNING IN NON-INTERACTIVE TEACHING FOR LOW-QUALITY EXPLANATIONS

Russ, H., Endres, T., Sibley, L., Flegr, S., Kuhn, J., Hoogerheide, V., Scheiter, K., & Lachner, A. (under review). Happy or redundant? Retrieval practice enhances lasting learning in non-interactive teaching for low-quality explanations.

The following manuscript has not yet been accepted or published. The version displayed here might not exactly replicate the final version published in the journal. It is not the copy of record.

5 Study 3: Happy or Redundant? Retrieval Practice Enhances Lasting Learning in Non-Interactive Teaching for Low-Quality Explanations

Abstract

Background: Supporting lasting learning is a central goal in education, particularly in schools where long-term retention is crucial for students' academic success. Generative learning helps construct meaningful mental representations, while retrieval practice supports consolidation and long-term retention. Combining both strategies may foster learning; however, research remains scarce.

Aims: We investigated how combining generative activities and retrieval practice affects students' conceptual knowledge and monitoring accuracy in authentic school settings for immediate and lasting learning. Additionally, we explored associations between generation, retrieval, and explanation quality (i.e., completeness, elaboration, correctness).

Sample: The classroom experiment was conducted in authentic physics lessons with $N = 344$ secondary students.

Methods: Students participated in a physics teaching unit about converging lenses and were randomly assigned within each class to one of four combinations of sequential learning activities crossing two factors (non-generation vs. generation; non-retrieval vs. retrieval). Conceptual knowledge and monitoring accuracy were measured immediately and after eight weeks.

Results: Conceptual knowledge and monitoring accuracy did not significantly differ between generation, retrieval, or their sequential combination at either test. However, exploratory analyses revealed that retrieval practice supported lasting learning only when prior generative processing was of low quality—that is, when students' explanations were incomplete, less elaborated, or inaccurate.

Conclusions: Combining generation and retrieval does not universally enhance learning. Its effectiveness appears to depend on the quality of prior generative processing. In this sense, retrieval practice may serve a happy role—compensating fragile knowledge—or a redundant one, offering no added benefit when prior understanding is already strong.

Keywords

Generative learning; learning by teaching; drawing; retrieval practice; lasting learning

5.1 Introduction

Lasting knowledge is pivotal for students' academic success and their ability to participate in today's knowledge society (OECD, 2023). However, recent large-scale assessments (e.g., Mullis et al., 2020; OECD, 2023) indicated that students often struggle to acquire and retain lasting knowledge. Generative activities are regarded as facilitating the meaningful construction of coherent mental representations. By engaging in these activities, students actively select, organize, and integrate new information, which fosters learning and metacomprehension (Fiorella, 2023b; Wittrock, 2010). Contrarily, retrieval practice, which involves retrieving previously learned contents, is considered essential for consolidation and lasting learning, that is, the long-term retention of mental representations (Karpicke, 2017). Given their distinct functions in learning, generative activities and retrieval practice may complement each other and mutually enhance learning outcomes. However, research on the combined effects of both learning activities is extremely limited (see Roelle, Endres, et al., 2023). Additionally, it is an open question how such combinations affect learning in authentic settings with school students regarding both immediate and lasting learning. Against this background, we conducted a classroom experiment with secondary students to investigate whether sequentially combining generative activities and retrieval practice can enhance (lasting) knowledge and monitoring accuracy.

5.2 Generative Learning Activities

Generative learning activities promote the active construction of knowledge and support meaningful learning (e.g., Fiorella, 2023b; Wittrock, 2010). Generative learning involves selecting, organizing, and integrating information to construct coherent mental representations (Wittrock, 2010).

One example is non-interactive teaching, where students explain contents to a fictitious peer (Lachner et al., 2022). The effectiveness of non-interactive teaching appears to be influenced by the quality of the explanations produced, including how thoroughly they cover key concepts (e.g., Jacob et al., 2020; blinded authors, 2025, under review) and the extent to which they incorporate detailed elaborations (e.g., Fiorella & Kuhlmann, 2020). Meanwhile, recent research has added visual components, acknowledging that teaching involves not only verbal explanations but also external representations (e.g., Fiorella & Kuhlmann, 2020; *blinded authors*, 2025, under review). Meta-analyses indicate small positive effects of non-interactive teaching on learning (e.g., Kobayashi, 2024: $g = 0.27$), and some evidence suggests benefits for monitoring accuracy. However, findings vary across and within studies, with outcomes ranging from positive to null or negative. Most studies relied on immediate or short-term delayed testing (e.g., one week), leaving the lasting effects of non-interactive teaching unclear. A rare example is the study by (*blinded authors*, accepted). In a classroom experiment embedded in regular physics lessons, secondary students ($N = 590$) either taught the previously learned contents to a fictitious peer, taught using a provided visualization, taught and drew, or restudied the material. Results showed that students in the generative conditions outperformed those who restudied the contents in an immediate posttest, but this advantage did not occur in delayed testing settings (eight weeks after the intervention). The inclusion of drawing in non-interactive teaching was particularly effective. These findings suggest that generative learning alone may not be sufficient for lasting learning, pointing to the potential value of additional consolidation activities such as retrieval practice.

5.3 Retrieval Practice

Retrieval practice involves recalling previously learned contents—typically through activities such as quizzes or tests—and is widely recognized as an effective strategy for promoting long-term retention (Yang et al., 2021). According to theoretical accounts of retrieval-based learning (Karpicke, 2017), the act of retrieval supports memory consolidation and enhances the durability of knowledge. Several hypotheses have been proposed to explain why retrieval practice may foster lasting learning. One prominent account suggests that retrieval strengthens memory traces by requiring effortful recall (Bjork & Bjork, 2011). Others emphasize elaborative retrieval (Carpenter, 2011) or context reinstatement (Rowland & DeLosh, 2014).

The benefits of retrieval practice, for instance, have been demonstrated in a classroom study by Roediger et al. (2011, Exp.1). The authors investigated the effects of repeated low-stakes quizzing during a sixth-grade social studies course. Students completed three multiple-choice quizzes (pretest, posttest, review test two days later) for half of the lesson contents, while the other half remained untested. Chapter exams two days later and a semester exam after one to two months revealed robust benefits for quizzed contents, with improvements in both free recall (immediate: η_p^2 = .64) and multiple-choice performance (immediate: η_p^2 = .73; delayed: η_p^2 = .45). These findings demonstrate that quizzes integrated into regular instruction can enhance both immediate and lasting learning outcomes.

There is also meta-analytic evidence supporting the benefits of retrieval practice in applied educational settings. Yang et al. (2021) synthesized findings from 222 classroom-based studies conducted in both school and university contexts and reported a medium overall effect of quizzing on students' achievement (*g* = 0.50). The effects were even larger when retrieval was combined with feedback ($g = 0.54$ vs. $g = 0.37$), which may help students fill knowledge gaps and reinforce memory traces.

5.4 Combining Generative Activities and Retrieval Practice

Considering the complementary functions of generative activities and retrieval practice, combining both activities may enhance students' learning. Roelle et al. (2023) emphasized that while generative activities produce meaningful mental representations, they often lack sufficient storage strength to ensure durable learning. Consequently, retrieval-based follow-up tasks may play a critical role in stabilizing these representations through consolidation. In the context of generative learning, retrieval practice could play a pivotal role by slowing students' forgetting of newly acquired knowledge (i.e., direct retrieval-practice effect, e.g., Rowland, 2014; Yang et al., 2021) and enhancing the effectiveness of subsequent relearning (i.e., indirect retrieval-practice effect, e.g., Endres et al., 2024). Following Roelle et al. (2023), how students engage in generative learning activities might affect the extent to which subsequent retrieval practice can be effective. For example, if generative activities are executed poorly, students may form only weak mental representations, providing a limited knowledge base for retrieval (see also Roelle, Froese, et al., 2022). Moreover, incorrect knowledge generated during these activities could later be retrieved and reinforced through retrieval practice, potentially consolidating misconceptions (e.g., Zhuang et al., 2022). However, the combination of a generative activity and retrieval practice might also induce a redundancy effect (Sweller et al., 2011), which occurs when the same information is processed through multiple sources that do not add value but instead impose unnecessary cognitive load. In such cases, additional learning activities—such as retrieval practice following a highly complete or accurate generative activity—may no longer contribute to learning and can even hinder it by burdening learners with redundant processing demands (see also Fyfe & Rittle-Johnson, 2016; Wagner et al., 2024).

Evidence on the interplay between generative activities and retrieval practice remains scarce, as research on these two approaches has largely been conducted independently (see Roelle, Endres, et al., 2023) or has focused on comparative horse-race studies (e.g., Jacob et al., 2020). Rather than identifying the most effective learning task, recent research has shifted toward investigating the benefits of combining generative learning and retrieval practice (see Roelle, Endres, et al., 2023). In this context, retrieval practice has been integrated into generative activities (e.g., Waldeyer et al., 2020), and generative learning has been incorporated into retrieval practice (e.g., Endres et al., 2024). However, findings suggest that combining both strategies does not always lead to superior learning outcomes compared to each approach individually (e.g., O'Day & Karpicke, 2021). Moreover, recent studies have explored the role of sequencing generative and retrieval tasks. Theoretical arguments suggest that a generation-before-retrieval approach may be beneficial as it enables learners to construct a well-organized mental representation first, which can later support more effective retrieval (Fiorella, 2023b; VanLehn, 1996). Conversely, retrieval-before-generation sequencing may strengthen memory traces, reduce cognitive load, and direct attention to key concepts, thereby facilitating subsequent generative learning (Roelle, Endres, et al., 2023; Roelle, Froese, et al., 2022; Rowland, 2014; Sweller et al., 2011). For example, Roelle et al. (2022) investigated the impact of sequencing retrieval practice and generative learning on knowledge acquisition. In an experiment, 158 university students first read a text about social attribution. After this initial study phase, they were randomly assigned to either a retrieval-before-generation or a generation-before-retrieval sequence. The retrieval task involved cued recall followed by feedback, which students processed by rating their answers against the provided correct responses. In the generative task, students were asked to generate their own examples of the accessible contents. The results showed that engaging in retrieval practice before generative learning led to greater retention gains (η_p^2 = .03, small effect) and reduced cognitive load during both tasks (η_p^2 = .07, medium effect) compared to the reverse order. However, as this is one of the first studies examining the sequencing of generative and retrieval activities, no definitive conclusions can yet be drawn regarding an optimal sequence, indicating that further research is needed.

Within the domain of learning by self-explaining, initial evidence suggests that especially the combination of retrieval practice and self-explaining promotes learning (Larsen et al., 2013). In this study, employing a within-subjects design, medical students took part in a teaching session covering four distinct topics. Over four weekly learning sessions, students performed one of four written learning tasks for each topic crossing two factors (restudy vs. retrieval, no-explaining vs. self-explaining). In a free-recall test after six months, the authors observed a main effect of retrieval (η_p^2 = .33, large effect) and self-explaining (η_p^2= .08, medium effect) and an interaction effect between self-explaining and retrieval (η_p^2 = .01, small effect). Additional pairwise comparisons revealed that the self-explaining condition yielded better learning performance when combined with retrieval rather than restudy (d = 0.70, medium effect). Moreover, there was a significant interaction among retrieval, self-explanation, and topic (η_p^2 = 0.06, small effect).

The findings, however, only apply to self-explanations and not to instructional explanations such as non-interactive teaching. While non-interactive teaching is a related but distinct explanation activity, it additionally requires students to adopt the audience perspective and adjust their explanations to enhance comprehensibility (see Lachner et al., 2021). Given that Larsen et al. (2013) conducted their study in a laboratory setting with university students and focused on self-explaining combined with retrieval, and considering that empirical evidence on the effectiveness of retrieval practice in problem-solving is mixed (Van Gog et al., 2015), it remains an open question whether these findings would generalize to the combination of generative activities and retrieval practice in authentic classroom settings with school students. Accordingly, it also remains unclear to what extent this approach fosters lasting learning in such authentic contexts.

5.5 The Present Study

Building on the previous considerations, the aim of this study was to investigate how the combination of generative activities (i.e., non-interactive teaching) and retrieval practice affect school students' learning in authentic classroom settings. We examined these effects on conceptual knowledge and monitoring accuracy in both immediate and lasting learning after eight weeks. Additionally, we explored whether the characteristics of the generative activities (i.e., completeness, elaboration, correctness) might be associated with the effectiveness of combined generation and retrieval. A classroom experiment in physics was conducted with seventh and eighth-grade students. We preregistered (https://aspredicted.org/wtpj-s7jd.pdf) the following hypotheses:

5.5.1 Generation Hypothesis (H1)

Grounded in the generative learning theory (Wittrock, 2010), we hypothesized that students engaging in a generative learning activity (i.e., non-interactive teaching) would outperform students in the non-generative activity (i.e., restudy) regarding their a) conceptual knowledge (Hoogerheide, Visee, et al., 2019; *blinded authors*, 2025, accepted) and b) monitoring accuracy (Jacob et al., 2020). In addition, we explored whether the effects were lasting, as assessed by a delayed test after eight weeks.

5.5.2 Retrieval Hypotheses (H2, H3[9])

Prior research suggests that retrieval practice fosters knowledge consolidation and lasting retention (Rowland, 2014). Accordingly, we hypothesized that students who retrieve learning contents (i.e., retrieval) would outperform students who only restudy the new contents

[9] Please note that the numbering of Hypotheses 3 and 4 has been reversed compared to the preregistration for improved readability. The content of the hypotheses remains unchanged.

(i.e., non-retrieval) regarding their a) conceptual knowledge and b) monitoring accuracy (H2). We also hypothesized that retrieval results in higher lasting learning than non-retrieval, as assessed by a delayed test after eight weeks (H3).

5.5.3 Interaction Hypothesis (H4)

Generative activities are regarded as contributing to elaboration processes, while retrieval practice should enhance consolidation (Roelle, Endres, et al., 2023; Roelle, Schweppe, Endres, Lachner, Aufschnaiter, et al., 2022). Therefore, we hypothesized that the combination of a generative activity and retrieval practice would yield synergistic effects on students' a) conceptual knowledge and b) monitoring accuracy. In addition, we explored whether the effects were lasting, as assessed by a delayed test after eight weeks.

5.5.4 Explorative Analyses on the Characteristics of Students' Explanations

Following Roelle et al. (2023), how students engage in the generative learning activity may shape the effectiveness of subsequent retrieval practice. Specifically, retrieval practice is likely to benefit learning only if a sufficiently accurate and elaborated mental representation has been constructed during the generative phase (see also Roelle, Froese, et al., 2022). Otherwise, retrieval may fail or even reinforce misconceptions (e.g., Roediger et al., 1996; Zhuang et al., 2022). Therefore, we explored whether the characteristics of students' explanations—completeness, elaboration, and correctness—moderate the effects of combined generation and retrieval.

5.6 Method

5.6.1 Participants and Design

A total of 372 seventh- and eighth-grade students from five secondary schools in southwestern Germany participated in the study. In total, 28 students were excluded because they missed the intervention, resulting in a final sample of $N = 344$. The sample exceeded the minimum of 206 participants determined by an a priori power analysis conducted in G*Power

(f = 0.25, α = .05, 1-β = .90, ANCOVA with one covariate), as dropout was lower than anticipated.

Students had a mean age of 12.58 years (*SD* = 0.63) and 49.20% identified as female. Regarding language background, 58.84% reported German as their native language, 16.08% were bilingual with German, and 25.08% had a different first language. Prior knowledge was relatively low (*M* = 8.91, *SD* = 3.83; maximum score: 30 points).

We applied a 2 × 2 between-subjects design crossing generative learning (non-generation vs. generation) and retrieval practice (non-retrieval vs. retrieval). Students either restudied the contents (non-generation: n = 168) or taught them to a fictitious peer (generation: n = 176), and subsequently either restudied (non-retrieval: n = 165) or completed a quiz (retrieval: n = 179). To avoid potential confoundings with the retrieval task, the generative task was conducted open-book, allowing access to the material to trace effects to a generative and not a retrieval function.

5.6.2 The Teaching Unit

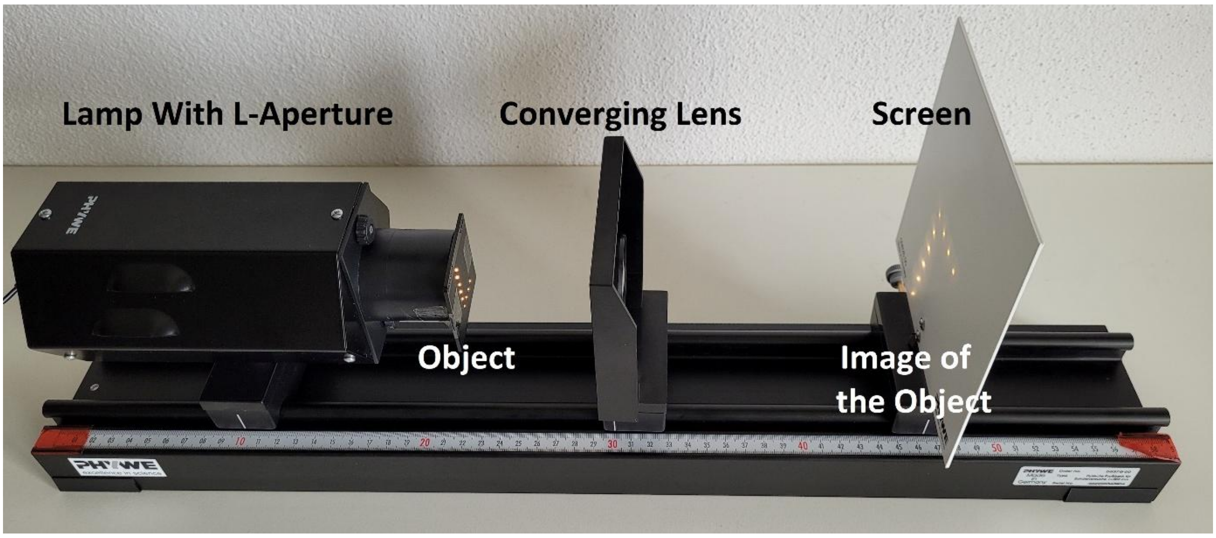
The physics teaching unit, aligned with the curriculum, covered converging lenses and image formation—a standard topic in geometrical optics for this age group. A certified physics teacher with 10 years of experience (first author) conducted all lessons and data collection. Instructional materials, including the introduction, overview sheet, and experimentation worksheet, were adapted from Flegr et al. (2023).

The lesson began with a teacher-led presentation using visual slides. Students activated prior knowledge by discussing everyday uses of lenses ("Where can you find lenses in your everyday life?"). Key concepts were introduced, including a definition, terminology, and an explanation of light refraction with examples. An overview sheet summarized core principles of converging lenses (see first worksheet by Flegr et al., 2023).

Following the introduction, students investigated how partially covering a lens and changing the object's distance affected image formation using an optical bench setup (Figure 16). A worksheet guided their inquiry: Students generated hypotheses, explained their reasoning, documented observations in a table, and compared them with the results.

Figure 16

Setup of the Students' Experiments on the Converging Lens and its Images

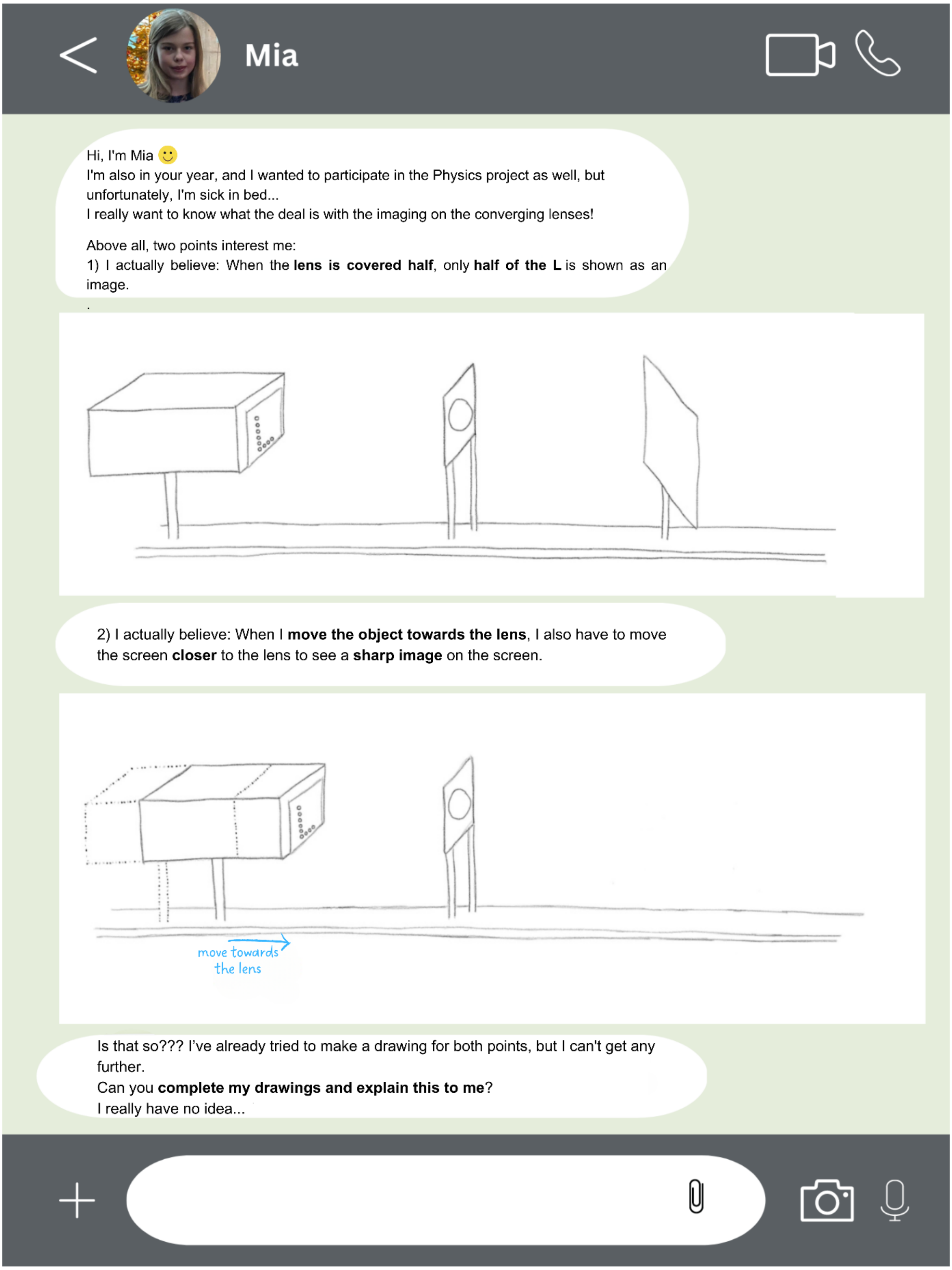


5.6.3 Generative Learning Activity

Students in the generative condition were instructed to teach the contents to the fictitious peer "Mia" by recording a voice message and drawing a visualization via a tablet-based chat. In contrast, students in the non-generative condition restudied the material. Both activities targeted two core objectives—effects of partial lens coverage and changes in object distance—and were framed around two related common misconceptions about image formation with a converging lens (see Figure 17; for task details, see *blinded authors*, 2025, accepted).

Figure 17

Mockup Messenger Chat of the Generative Learning Activity



Note. The mock-up chat displays a message from the fictitious peer Mia, accompanied by a profile picture. Students in the generative condition were able to record a voice message and use the two provided template pictures for drawing. Translated from German

5.6.4 Retrieval Practice

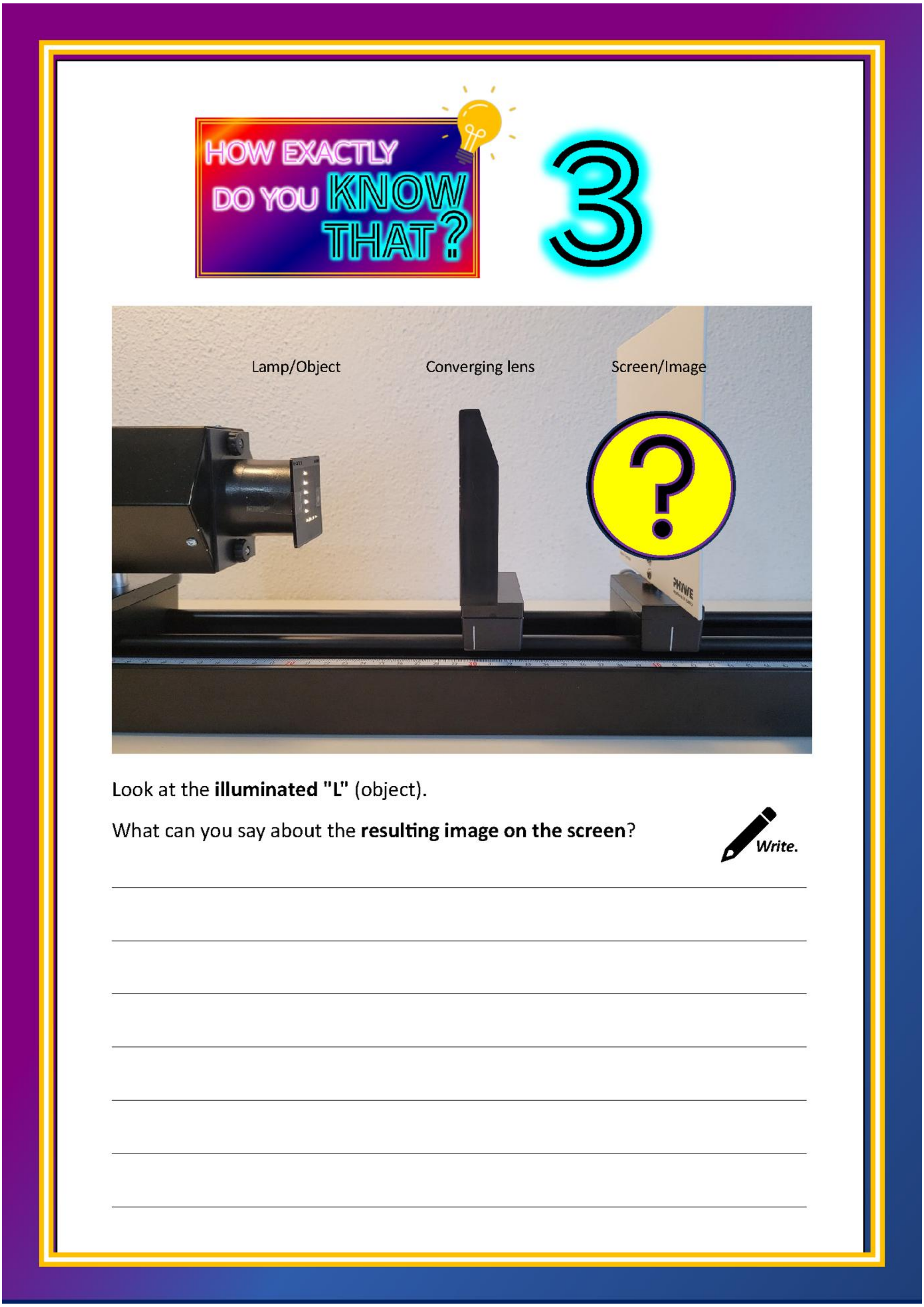
In addition to the generative activity, students were randomly assigned to either a non-retrieval or retrieval condition, resulting in a 2 (non-generation vs. generation) × 2 (non-retrieval vs. retrieval) between-subjects design.

Students in the retrieval condition completed a paper-based quiz titled "How exactly do you know that?" (available on OSF) without access to the learning material (closed-book approach, see Roelle, Endres, et al., 2023; for similar implementation, see McDermott et al., 2014), following typical classroom-based retrieval formats (see Yang et al., 2021). The quiz included five questions addressing core concepts of image formation with converging lenses (for related core concepts, see Wörner et al., 2022), covering the same two learning objectives as the previous contents: the partly covered lens and the changing distance between the object and the lens, along with the resulting image formation. The first three questions were open-ended (e.g., "How does a converging lens image an object?"; see Figure 18), the last two were multiple-choice.

The sequence of questions followed a progression from abstract to concrete. This design encouraged a broad retrieval of knowledge by addressing many concepts at the beginning and gradually narrowing to focus on fewer, more specific concepts. It transitioned from recognition to recall and from minimal textual retrieval cues alone to a more extensive and detailed combination of textual and pictorial retrieval cues (see Endres et al., 2024; Yang et al., 2021). After completing the quiz, students received feedback by presenting the correct marked answer or in terms of the open-ended questions in form of correct idea units necessary for a complete answer (for similar approach, see e.g., Roelle, Froese, et al., 2022).

Figure 18

Example Question From the Retrieval Quiz



5.6.5 Measures

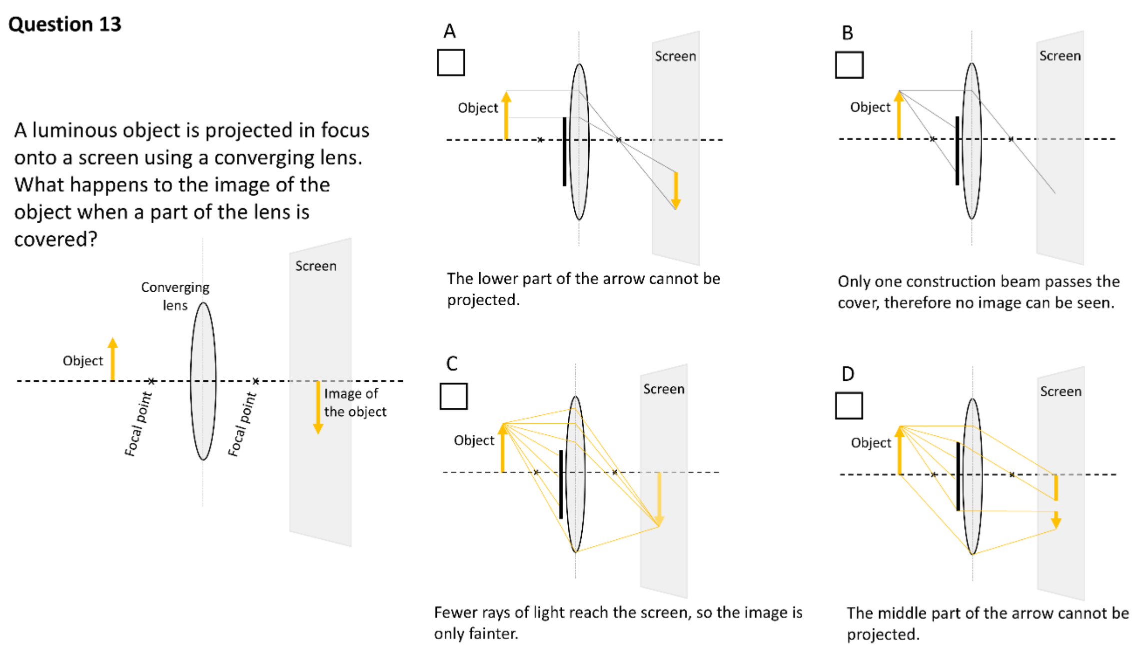
Data were collected on paper. As part of a larger research project, this study focuses on the variables relevant to its objectives (see Appendix A for a comprehensive list).

5.6.5.1 Conceptual Knowledge

To assess students' conceptual knowledge at pre-, post-, and delayed tests, we employed the validated ROC-CI test by Wörner et al. (2022), which included 15 multiple-choice items targeting understanding of converging lenses, with distractors reflecting common misconceptions (see Figure 19). Correct answers scored two points, partially correct ones one point (max = 30). To reduce recognition bias, answer options were randomized across all test phases. Two independent raters evaluated 20% of the responses ($ICC_{2,1} = 0.99$); one rater scored the remainder. The reliability of the test was satisfying ($\omega_t = 0.68$).

Figure 19

Sample Item From the ROC-CI Conceptual Knowledge Test on Converging Lenses



Note. From Wörner et al. (2022).

5.6.5.2 Monitoring Accuracy

To assess students' metacognitive monitoring accuracy, we asked the students to estimate their expected performance prior to each of the knowledge tests (pre-, post-, and delayed): "In the following you will answer 15 questions about the topic 'Imaging by converging lenses.' You can get two points per question. In total, you can score 30 points. How many points do you think you will get?" (e.g., Baars et al., 2017). Students responded on a scale from 0 to 30, consistent with previous research (e.g., Jacob et al., 2022). Monitoring accuracy was calculated using a bias score[10] (Schraw, 2009) by determining the difference between the predicted and actual scores (i.e., $X_{Judgment} - X_{Performance}$), separately for each test time point. Negative values (minimum: -30) indicate underestimation, positive values (maximum: 30) overestimation, and zero perfect accuracy.

5.6.5.3 Additional Control Measures

Students' Prerequisites. To account for potential differences between experimental conditions, we assessed students' interest in physics ($\omega_t = 0.85$), physics work ethic ($\omega_t = 0.77$), academic self-concept in physics ($\omega_t = 0.87$), and ICT interest ($\omega_t = 0.69$). All constructs were measured with four items on a Likert scale from one "I completely disagree" to four "I completely agree" and adapted from Mang et al. (2018, 2019).

Cognitive Load. Students rated cognitive load at two time points (after the teaching unit and the interventions), assessing active load (i.e., invested effort) and passive load (i.e., experienced load). Both aspects were assessed on a nine-point Likert scale from "not strenuous at all" to "very strenuous" (Klepsch & Seufert, 2021; Paas, 1992).

[10] In our preregistration, we inadvertently used the term 'absolute' in the detailed description of the bias measure for monitoring accuracy. To maintain transparency and uphold scientific rigor, we have acknowledged this wording error. This aligns with Lakens' (2024) recommendation that addressing preregistration errors openly enhances the validity of the analysis.

Affect. After the teaching unit, students rated their arousal and mood (Betella & Verschure, 2016) on two Likert scales from one "sleepy/bored" to nine "wide awake/focused" (arousal) and from one "sad/in a bad mood" to nine "happy/in a good mood" (mood).

Teaching Quality. To ensure high implementation fidelity, students evaluated the teaching quality immediately after the teaching unit, rating cognitive activation ($\omega_t = 0.78$), student disturbances ($\omega_t = 0.85$), teacher monitoring ($\omega_t = 0.85$), and teacher support ($\omega_t = 0.84$). Scales were adapted from Fauth et al. (2014), Baumert et al. (2012), and Mang et al. (2019) and used four-point Likert scales from one "I completely disagree" to four "I completely agree".

5.6.6 Procedure

The study received approval from the university's ethics committee and the local Ministry of Education. Participation was voluntary, and data was collected only from students with written consent from their legal guardians. A summary of the procedure is presented in Figure 20.

About a week before the teaching unit, the first author visited each school class to introduce the study and administer the pretest (approx. 30 minutes), assessing students' demographics, prerequisites, prior conceptual knowledge, and monitoring accuracy.

One week later, the first author, a certified physics teacher with 10 years of experience, conducted the main study part. Students were introduced to converging lenses (15 minutes), received an overview sheet, and performed experiments guided by a worksheet (30 minutes). During this phase, student-teacher interactions resembled typical physics lessons, allowing for guidance when needed. The session concluded with a review of experimental results. Afterwards, students completed a questionnaire on cognitive load, affect, and teaching quality and were then randomly assigned to one of four conditions: non-generation + non-retrieval, non-generation + retrieval, generation + non-retrieval, or generation + retrieval.

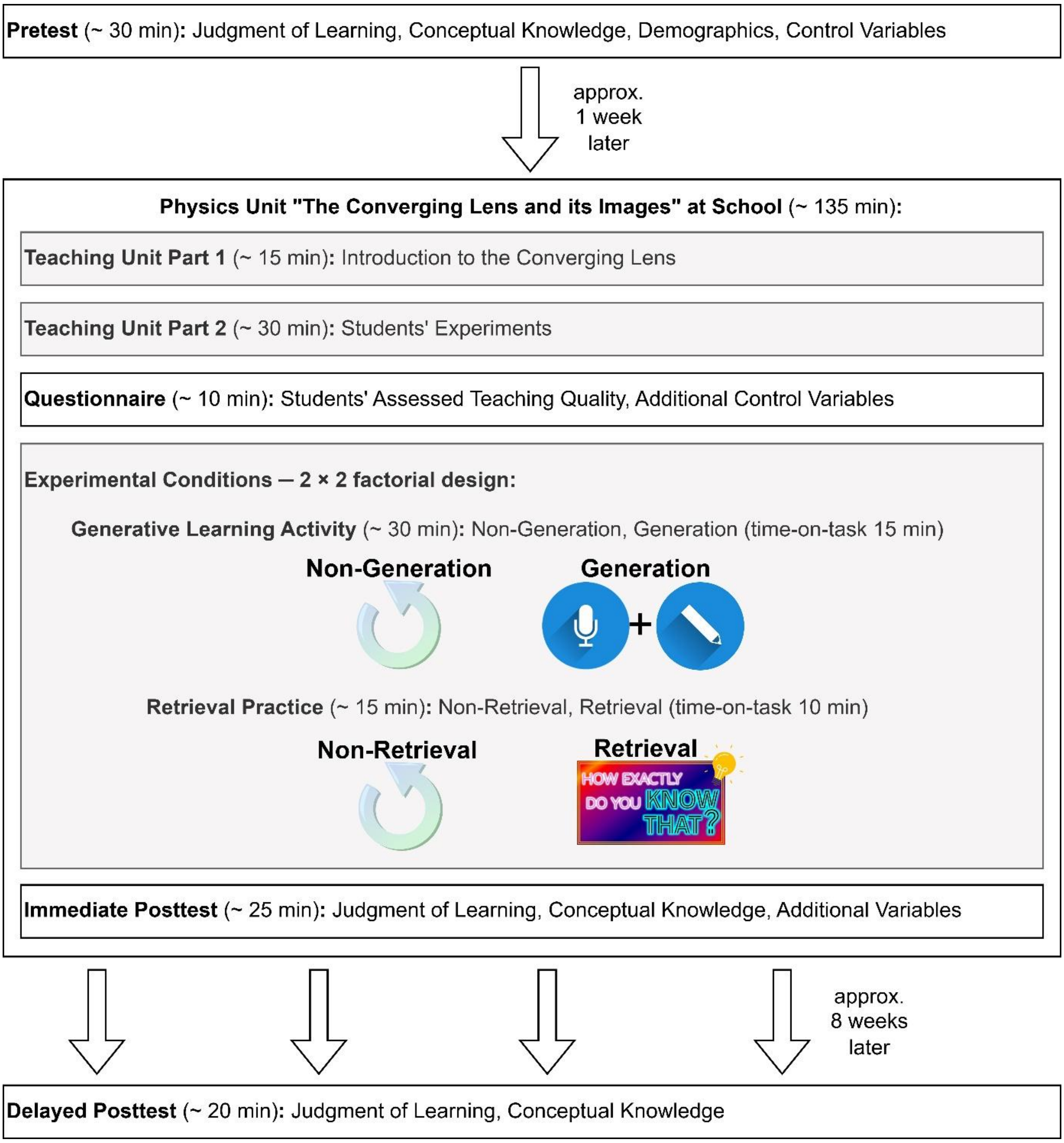
During the learning activity phase (time-on-task 15 minutes), students either taught the contents to a fictitious peer (generation) or restudied the material (non-generation). Groups were separated, and students in the generative condition wore ear protectors. All had access to the material to minimize potential retrieval confounds (Roelle, Endres, et al., 2023; Sibley et al., 2022).

During the subsequent retrieval practice phase (time-on-task 10 minutes), students either completed a paper-based retrieval quiz (retrieval) or restudied the material (non-retrieval). To prevent distractions, groups were separated. Quiz takers received a prompt after two minutes to proceed to the next question and were not allowed to skip questions or revise answers (for similar approach, see McDermott et al., 2014). Then, students received feedback on their retrieval task.

Finally, all students took the immediate posttest (approx. 20 minutes). The delayed posttest followed approximately eight weeks later.

Figure 20

Design and Procedure of the Study



5.6.7 Analysis of Students' Explanations, Drawings, and Restudy Notes

Explanations, drawings, and restudy notes were coded for completeness, elaboration, and correctness as indicators of underlying cognitive processes (see also Fiorella & Kuhlmann, 2020; Hoogerheide, Renkl, et al., 2019; Schmidgall et al., 2019). Interrater reliabilities of 20%

of the data were excellent ($0.92 \leq ICC_{2,1} \leq 1.00$). The detailed analysis scheme is accessible at: https://osf.io/cr5yn/?view_only=453ffd61cdd0444c9ba108296b21c8c3.

5.6.8 Analysis of Students' Retrieval Quiz

We analyzed how many core concepts students recalled during the retrieval task. A maximum of 13 core concepts could be retrieved in total (for core concepts, see Wörner et al., 2022). Interrater reliability for 20% of the coding was excellent ($ICC_{2,1}$ = 0.92−1.00). The detailed coding scheme is available at:

https://osf.io/cr5yn/?view_only=453ffd61cdd0444c9ba108296b21c8c3.

5.6.9 Data Analyses

To test our preregistered hypotheses, we used structural equation modeling (SEM) with the lavaan package (Rosseel, 2012) in R (version 4.4.3; R Core Team, 2025), controlling for students' prior conceptual knowledge or prior monitoring accuracy. We also explored whether completeness, elaboration, and correctness of explanations or restudy notes moderated effects within SEM (Appendix B).

Although our preregistration specified multiple imputation, we chose Full Information Maximum Likelihood (FIML) because it uses all available data to estimate model parameters without generating multiple datasets, reducing bias and preserving power, especially in small to moderate samples (Enders, 2022; Newman, 2003). Compared to multiple imputation, FIML more directly models missingness and yields more precise and efficient estimates with lower computational complexity (Graham, 2009).

Accordingly, we used SEM instead of our preregistered ANCOVAs. This allowed us to model multiple dependencies simultaneously, account for measurement error, and address the nested data structure (students nested within classes and schools) via multilevel modeling with cluster-robust standard errors. SEM also enabled an integrated analysis of direct and interaction effects while handling missing data robustly (Little, 2024; J. Wang & Wang, 2020).

Our decision to deviate from the preregistered 2×2 ANCOVAs and multiple imputation aligns with recommendations by Lakens (2024) to prioritize methodological rigor when alternative approaches provide more precise and robust inferences.

5.7 Results

The preregistration, dataset, analyses, and supplementary material can be accessed at: https://osf.io/cr5yn/?view_only=453ffd61cdd0444c9ba108296b21c8c3. Effect sizes were reported using Cohen's d, partial η_p^2, and φ, with thresholds defined as follows: small effects (d = .20, η_p^2 = .01, φ = .10), medium effects (d = .50, η_p^2 = .06, φ = .30), and large effects (d = .80, η_p^2 = .14, φ = .50; see Cohen, 2013). The significance level was set at α = .05.

5.7.1 Preliminary Analysis

Our boxplot analysis showed no indication of outliers in the data. Further preliminary analyses revealed that conditions (non-generation + non-retrieval, non-generation + retrieval, generation + non-retrieval, generation + retrieval) did not differ regarding gender, χ^2 (6, 344) = 4.58, p = .598, φ = .12, or first language, χ^2 (6, 344) = 10.45, p = .107, φ = .18. A MANOVA indicated no significant differences between the conditions in terms of students' age, interest in physics, physics work ethic, academic self-concept in physics, ICT interest, prior conceptual knowledge, and prior monitoring accuracy, F(3,344) = 0.82, p = .700, φ = .02. A second MANOVA revealed no significant differences in students' cognitive load, students' affect, and perceived teaching quality regarding the teaching unit, F(3,344) = 0.98, p = .486, φ = .02. Appendix B provides the mean scores and standard deviations for each condition, while correlations between the variables are presented in Appendix C.

As an implementation check, we compared the characteristics of restudy notes and generated explanations from the generative learning activity phase. Explanations were more complete (β = .20, p = .030), more elaborated (β = .79, p < .001), but not more correct (β = –

.22, p = .167) than restudy notes. Students in the retrieval practice condition retrieved an average of 5.59 out of 13 possible core concepts during the quiz (SD = 0.91).

After the interventions, students in the retrieval condition reported higher actively invested effort than those in the non-retrieval condition (β = .35, p = .028), while no significant differences emerged between non-generation and generation (β = .27, p = .150). Post-hoc tests revealed that students who both generated and retrieved reported more effort than those who did neither (β = .47, p = .019). For passively experienced load, both generation (β = .45, p = .002) and retrieval (β = .44, p = .026) led to higher ratings. Students who engaged in generation and retrieval reported more passive load than those who did neither (β = .60, p = .001), as did those who only retrieved (β = .42, p = .038) or only generated (β = .47, p = .017).

5.7.2 Does Engaging in a Generative Learning Activity Enhance Immediate and Lasting Learning?

In contrast to our generation hypothesis (H1), generative learning did not lead to higher conceptual knowledge or improved monitoring accuracy compared to non-generative learning. No main effect of generation was detected at either the immediate or delayed posttest after eight weeks (.093 ≤ p ≤ .969; see Appendix D).

5.7.3 Does Retrieval Practice Enhance Immediate and Lasting Learning?

Contrary to our retrieval hypotheses (H2, H3), retrieval practice did not significantly improve conceptual knowledge or monitoring accuracy. No main effect of retrieval was detected at either the immediate or delayed posttest (.307 ≤ p ≤ .783; see Appendix D).

5.7.4 Does the Sequential Combination of a Generative Learning Activity and Retrieval Practice Yield Synergistic Effects on Immediate and Lasting Learning?

Against our interaction hypothesis (H4), sequentially combining generation and retrieval did not result in synergistic effects on conceptual knowledge or monitoring accuracy.

No interaction effect was detected at either the immediate or delayed posttest ($.268 \leq p \leq .726$; see Appendix D).

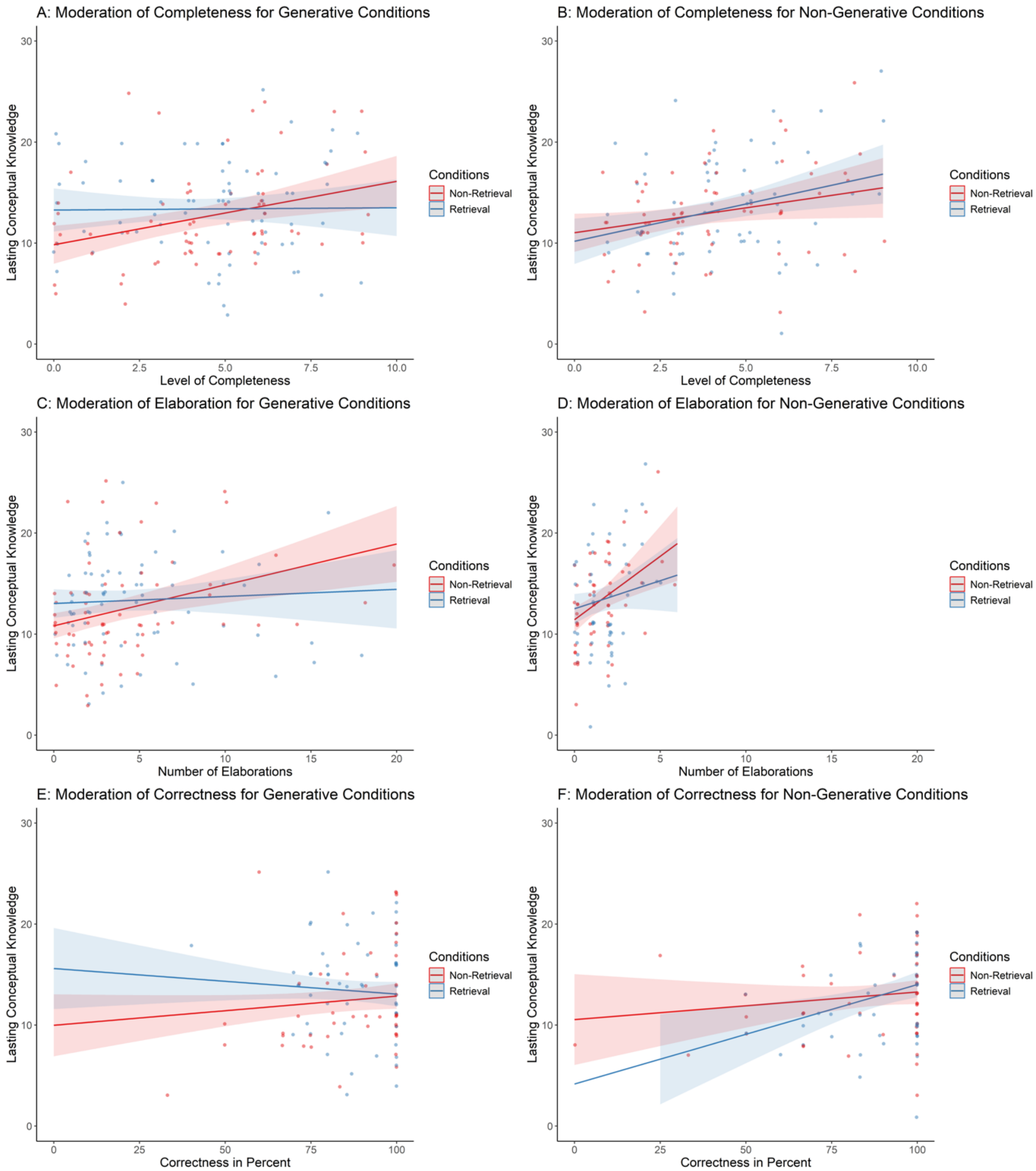
5.7.5 Does the Effectiveness of Retrieval Practice Depend on the Quality of Students' Generative Processing?

The effectiveness of sequentially combining generation and retrieval may depend on the quality of the mental representations constructed during the initial phase (Roelle, Endres, et al., 2023). Retrieval is only likely to be effective if it builds on sufficiently accurate and elaborated mental representations (Roelle, Endres, et al., 2023), which might explain why the sequential combination of generation and retrieval did not yield overall synergistic effects. Therefore, we explored whether the completeness, elaboration, and correctness of students' restudy notes or explanations moderated the effect of retrieval practice on conceptual knowledge. Separate models were estimated for each characteristic to avoid multicollinearity and preserve statistical power (Aguinis, 1995; for correlations, see Appendix C; for full model results, see Appendix E).

Across all three characteristics, the three-way interactions between generation, retrieval, and the respective characteristic were not significant for immediate learning ($.201 \leq p \leq .738$) but reached significance for lasting conceptual knowledge after eight weeks ($.002 \leq p \leq .007$). Interaction plots (see Figure 21) revealed a consistent pattern: For students in the generative conditions, those with low to moderate levels of completeness, elaboration, or correctness benefited more from retrieval practice than those with high levels. In contrast, within the non-generative conditions, retrieval practice did not yield consistent learning benefits—regardless of the level of completeness, elaboration, or correctness.

Figure 21

Moderation of Completeness, Elaboration, and Correctness on Lasting Conceptual Knowledge in Generative and Non-Generative Conditions



5.8 Discussion

In this experimental classroom study in authentic physics lessons, we aimed to investigate whether the combination of a generative activity and retrieval practice enhances conceptual knowledge and monitoring accuracy, both immediately and after a delay of eight weeks. Additionally, we explored the quality of generative processes as potential moderators of the effectiveness of subsequent retrieval practice.

Despite the rationale for combining generative activities and retrieval practice, the current study did not reveal significant main effects of generation, retrieval, or a synergistic effect of their sequential combination on students' conceptual knowledge or monitoring accuracy, neither immediately nor after a delay of eight weeks (H1-4). These findings diverge from prior research that has demonstrated individual benefits of generative learning and retrieval practice in more controlled environments (see Adesope et al., 2017; Kobayashi, 2024), as well as in authentic classroom contexts (e.g., *blinded authors*, 2025, accepted). Laboratory settings typically differ from authentic classrooms in terms of study populations and contextual factors such as learner diversity, environmental distractions, and social dynamics (e.g., Dinsmore & Alexander, 2012), and even classroom environments may vary considerably in these respects. Moreover, the impact of retrieval practice may depend on how students engaged in the generative activity. If this activity failed to produce a sufficiently elaborated and well-integrated knowledge base—for example, due to superficial processing or misconceptions—retrieval may not have had an impact on learning outcomes. In other words, the success of retrieval practice may have depended not only on storage strength (Roediger & Butler, 2011) but also on the quality of what students retrieved (see also Roelle, Endres, et al., 2023).

Therefore, we explored whether the effectiveness of retrieval practice depended on how well students had engaged in the generative activity, based on the quality of the generated explanations (i.e., completeness, elaboration, and correctness). None of the three-way interactions were significant for immediate learning. However, after eight weeks, all three characteristics consistently influenced the effects: Retrieval practice benefited students whose generative processing had been of rather low quality, whereas those with high-quality explanations showed no additional gains.

Retrieval practice following a generative activity appeared to be most effective when the generative processing had not yet resulted in sufficiently coherent knowledge. Rather than merely reinforcing existing representations, as proposed by Roelle et al. (2023), retrieval may have served as a form of productive struggle: It required students to reconstruct and consolidate fragmented representations, thereby enhancing long-term retention. This interpretation may reflect the notion of desirable difficulties (Bjork & Bjork, 2011), which suggests that effortful—but ultimately successful—processing promotes lasting learning. Supporting this view, students who both generated and retrieved, reported higher levels of invested effort than those who did neither. In this happy sense, retrieval practice may have helped students revisit, refine, and stabilize initially fragile mental representations resulting from the generative activity, turning them into more durable knowledge.

The lack of additional benefits among students who had already produced highly complete, elaborated, and correct explanations may reflect a redundancy effect: When generative processing has already led to sufficiently coherent knowledge, additional retrieval practice may no longer enhance learning. According to cognitive load theory, reprocessing well-integrated information can impose unnecessary demands, as learners must monitor and sort through redundant input—thereby increasing extraneous cognitive load and impairing learning efficiency (Sweller et al., 2011). Supporting this view, both the generative and retrieval tasks in our study targeted the same learning objectives. Moreover, students who both generated and retrieved reported higher experienced cognitive load than those who did neither—and also more than those who only generated. This pattern aligns with prior research showing redundancy effects in feedback settings (e.g., Fyfe & Rittle-Johnson, 2016; Wagner et al., 2024) and suggests that similar effects may also arise in the context of subsequent generative and retrieval activities.

Taken together, our findings suggest that retrieval practice following a generative activity can be either *happy* or *redundant*, depending on the quality of students' generative processing—either compensating for initially fragile representations or yielding no observable learning gains when prior processing has already produced coherent knowledge. Perhaps it would have been helpful if students had engaged in repeated cycles of generation and retrieval across multiple sessions, as sustained engagement over time may be more effective in fostering learning (Rawson & Dunlosky, 2022; see also Larsen et al., 2013). It is also important to consider that instructional designs involving generation and retrieval may not benefit all students equally (see also Snow, 1991). Future research should therefore explore how to best tailor generative and retrieval-based activities to students' needs to maximize their potential in authentic classroom settings.

5.8.1 Study Limitations and Future Research

The present study was conducted in authentic classroom settings, which enhances ecological validity but also introduces limitations that may constrain the generalizability of the findings. We made considerable efforts to ensure consistent instructional quality and standardized procedures across all classes; for instance, all lessons and study phases were conducted personally by the first author, who is both a physics teacher with ten years of experience and a researcher. While this ensured fidelity of implementation, it remains unclear whether the observed patterns would replicate with different teachers or under more typical school conditions. Moreover, we focused on physics in seventh and eighth grade classes, and prior research has shown that subject matter characteristics and students' developmental differences may interact with learning activities (see, e.g., Brod, 2021; Sibley et al., 2024). Future research should therefore examine whether similar effects emerge when interventions are implemented by the students' regular teachers, and whether they generalize to other disciplines such as mathematics or language learning and to students across different age groups.

The study employed a generation-before-retrieval sequence, following theoretical assumptions about mental model construction and knowledge consolidation (Roelle, Endres, et al., 2023). However, recent work suggests that retrieval-before-generation may reduce cognitive load and foster attention to key concepts, thereby supporting subsequent generative processing (Roelle, Froese, et al., 2022). Future studies should systematically compare different sequencing approaches (generation-before-retrieval vs. retrieval-before-generation) to determine their relative effectiveness under diverse instructional conditions.

Finally, the generative learning activity in this study involved a combination of non-interactive teaching and drawing. While prior research has demonstrated the benefits of this combination (e.g., Fiorella & Kuhlmann, 2020; *blinded authors*, b), other generative activities such as enacting have also shown promise in fostering learning (Novack & Goldin-Meadow, 2015). Future research could explore the potential synergistic effects of alternative generative activities in combination with retrieval practice to identify which combinations are effective for supporting learning in authentic settings.

5.9 Conclusion

In our classroom study with an immediate and a delayed posttest eight weeks later, we investigated the effects of combining a generative activity with subsequent retrieval practice in authentic physics lessons. While neither generation, retrieval, nor their sequential combination yielded significant main effects on students' conceptual knowledge or monitoring accuracy, exploratory analyses revealed that the quality of students' generative processing played a moderating role. Retrieval practice particularly supported lasting learning when students' prior explanations were incomplete, less elaborated, or contained inaccuracies, but did not enhance learning when explanations were already of high quality. These findings suggest that the effectiveness of retrieval practice is not universal, but contingent on the cognitive conditions it builds upon. In this sense, whether the combination of generation and subsequent retrieval turns out to be *happy* or *redundant* in classroom practice may ultimately depend on the quality of students' prior generative engagement. These insights highlight the importance of considering students' processing and the fact that instructional effectiveness is not one-size-fits-all.

6

GENERAL DISCUSSION

6 General Discussion

This final chapter summarizes the central findings of the three empirical classroom studies and situates them within prior research and the Offer–Use Model of Teaching for Students' Non-Interactive Teaching (Figure 22). The goal is to reflect on how the results contribute to answering the overarching research question and what implications they hold for refining the Offer–Use Model of Teaching for Students' Non-Interactive Teaching, understanding learning processes in authentic classroom settings, and informing educational research and practice more broadly, including contexts that serve students from disadvantaged backgrounds and schools in socially deprived areas.

6.1 Discussion of the Results

The following section begins by summarizing the main empirical findings across the three studies. These findings are then interpreted in light of the Offer–Use Model of Teaching for Students' Non-Interactive Teaching and discussed in relation to prior research on generative learning and retrieval practice. Given that all studies were conducted in inquiry-based, authentic classroom settings, special attention is given to interpreting the results within their educational context. Finally, implications for expanding and refining the Offer–Use Model of Teaching for Students' Non-Interactive Teaching are derived.

6.1.1 Summary of the Results

This dissertation investigated how the generative learning activity of non-interactive teaching can be optimized regarding students' conceptual knowledge and monitoring accuracy in inquiry-based authentic science lessons, in terms of immediate and lasting learning. Three empirical classroom studies tested theoretically and empirically grounded instructional modifications of non-interactive teaching, focusing on drawing, distribution, and retrieval practice.

Study 1 investigated whether non-interactive teaching can be enhanced by adding drawing, with regard to students' conceptual knowledge and monitoring accuracy. Results showed that non-interactive teaching had an effect on immediate conceptual knowledge, which was further improved when students processed a provided picture and improved even more when they generated their own drawing. No significant effects were found on lasting learning outcomes or monitoring accuracy. Exploratory mediation analyses suggested that task-specific motivation and the completeness of students' explanations partially accounted for the observed learning effects.

Study 2 investigated whether non-interactive teaching can be enhanced by distributing the learning activity throughout the study phase, with respect to students' conceptual understanding and monitoring accuracy. The results showed no overall effect of distribution. However, exploratory analyses of potential interactions between non-interactive teaching and distribution suggested that non-interactive teaching was effective only when implemented once after the study phase, but not when distributed throughout. Exploratory moderation analyses indicated that the effectiveness of non-interactive teaching on immediate conceptual knowledge was associated in part with individual student characteristics such as academic self-concept and work ethic—however, only in the after-study condition, not in the distributed one. In the after-study condition, students with higher levels of academic self-concept or work ethic benefited more from the generative conditions compared to restudying. For immediate monitoring accuracy, students in the restudy condition numerically overestimated their knowledge, while those in the generative conditions significantly underestimated their performance. No effects emerged for lasting outcomes. Moreover, adding drawing to non-interactive teaching did not yield further benefits beyond non-interactive teaching alone.

Study 3 investigated whether non-interactive teaching can be enhanced by sequentially combining it with retrieval practice, with respect to students' conceptual understanding and monitoring accuracy. Contrary to the hypotheses, no main effects of generation, retrieval, or their sequential combination were observed on either immediate or lasting conceptual knowledge or monitoring accuracy. However, exploratory moderation analyses revealed that the effectiveness of retrieval practice depended on the quality of prior generative processing. Specifically, students with lower levels of completeness, elaboration, or correctness in their generative activity benefited more from subsequent retrieval practice than those with high-quality generative output. This pattern emerged consistently for lasting conceptual knowledge, but not for immediate outcomes.

6.1.2 Situating the Results Within Prior Research and the Offer-Use Model of Teaching for Students' Non-Interactive Teaching

The results of Study 1 showed that non-interactive teaching could be enhanced by combining it with drawing; this combination improved students' immediate conceptual knowledge. These results align with previous findings from laboratory studies with university students, suggesting that adding visual generative activities can deepen understanding (Fiorella, 2023a; Fiorella & Kuhlmann, 2020). However, they also contrast with findings by Fiorella (2023a), who reported that non-interactive teaching with provided visualizations resulted in higher drawing test performance than the combination of teaching and drawing. It should be noted, however, that Fiorella's study was not conducted in an authentic, inquiry-based classroom setting, as was the case in the present first study. In addition, no specific drawing test was administered in our study, limiting direct comparability of this outcome. From the perspective of the Offer–Use Model of Teaching for Students' Non-Interactive Teaching (see Figure 22), the additional drawing component represents a more elaborated offer, which appears to facilitate deeper cognitive engagement during non-interactive teaching. The partial mediation by motivation and explanation quality highlights that student-related factors and active use of the offer are essential for its effectiveness. These findings support the model's assumption that teaching success results from a reciprocal relationship between the offer and the way it is taken up by the student, rather than from the offer alone (see also Vieluf et al., 2020).

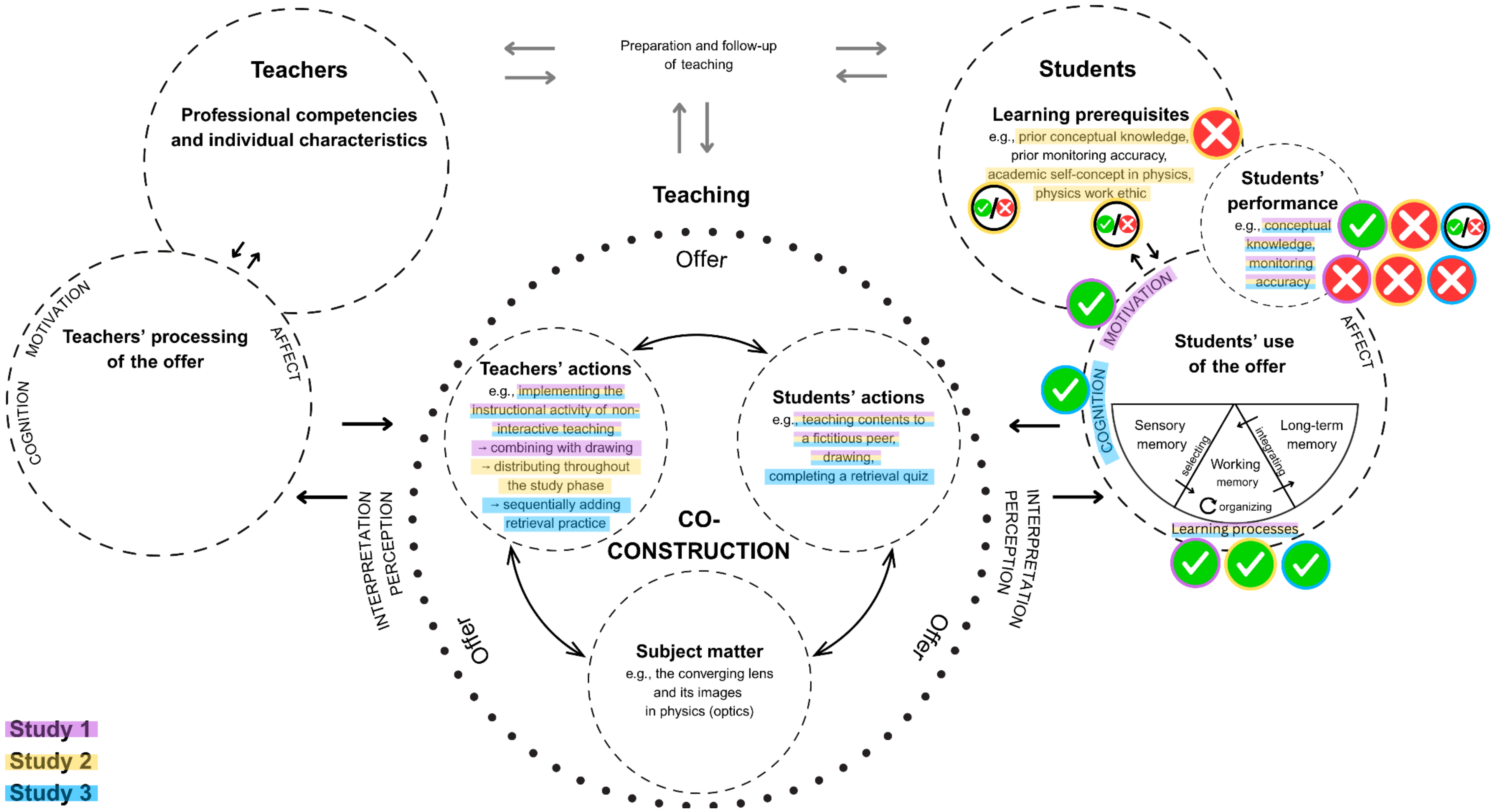
Study 2 investigated whether non-interactive teaching can be enhanced by distributing the activity across multiple points within the study phase. The findings showed no overall advantage of distribution for conceptual understanding or monitoring accuracy. Exploratory analyses indicated that non-interactive teaching was only effective when implemented once after the study phase. These findings contrast with earlier results by Lachner et al. (2020), who reported benefits of interpolating non-interactive teaching once during the study phase. However, that study was conducted in a laboratory setting with university students, using only one teaching opportunity, and did not investigate distribution across multiple points in time. The present results also contrast with assumptions drawn from related interpolated testing research (Pan et al., 2024), which has highlighted the benefits of inserting testing at multiple points. In the present second study, however, such benefits did not generalize to generative activities like non-interactive teaching when implemented repeatedly within the complexity of authentic inquiry-based classroom settings. From the perspective of the Offer–Use Model of Teaching for Students' Non-Interactive Teaching (see Figure 22), these findings suggest that simply increasing the frequency of an offer does not ensure its effective use. The modification implemented in Study 2 may have unintentionally altered other aspects of teaching. For example, it may have disrupted the coherence of the inquiry sequence or increased the cognitive demands of the task, which could have affected how the offer was taken up by students. The observed moderation effects of academic self-concept in physics and physics work ethic—limited to the after-study condition—underscore the model's assumption that teaching success emerges from the interaction between teacher, student, and subject matter, and is shaped by students' learning prerequisites. This also illustrates the complexity of teaching processes in the classroom (see also Vieluf et al., 2020). The fact that students' prior knowledge did not moderate the effects of non-interactive teaching (for similar results, see Jacob et al., 2022; J. Richter et al., 2022) may be due to the adaptable nature of this generative learning activity (Brod, 2024). Importantly, this presents a chance to enable learning success among students with limited prior knowledge—an aspect that may be particularly beneficial for those from disadvantaged backgrounds (see also Klein, 2017). Accordingly, non-interactive teaching represents an offer that enables students to process contents in ways that align with their individual prior knowledge—thereby supporting a form of domain-specific co-construction within the teaching context. However, this possible mechanism—namely the adaptive processing of teaching offers based on prior knowledge—is not explicitly represented within the current formulation of the Offer–Use Model of Teaching for Students' Non-Interactive Teaching (Figure 22).

Study 3 investigated whether non-interactive teaching can be enhanced by sequentially adding retrieval practice. Although no overall benefits emerged across all students, exploratory analyses revealed a targeted effect on lasting conceptual knowledge: students who had previously produced low-quality generative explanations benefited substantially from the subsequent retrieval activity. The present results contrast with theoretical assumptions and previous research reporting overall benefits of sequentially combining generative and retrieval-based learning activities (e.g., Larsen et al., 2013; Obergassel et al., 2025; see also Roelle, Endres, et al., 2023). In contrast to these laboratory-based studies, the present study implemented a sequential combination of non-interactive teaching and retrieval practice in an inquiry-based authentic school setting. Exploratory analyses indicated that this combination was particularly beneficial for students who had previously produced low-quality generative explanations, suggesting that the success of such combinations may hinge on the quality of prior processing and possibly on additional factors inherent to complex classroom environments (see also Figure 22). From the perspective of the Offer–Use Model of Teaching for Students' Non-Interactive Teaching (see Figure 22), this pattern suggests that teaching offers may interact dynamically across time. The effectiveness of retrieval depended not solely on its design as an offer, but on the quality of preceding student processing—that is, how deeply the initial offer (i.e., non-interactive teaching) was used. This finding is particularly relevant in light of the SOI model of generative learning (Fiorella & Mayer, 2016), which conceptualizes meaningful learning as the integration of selected and organized information into existing knowledge structures. In the adapted model (Figure 22), these internal learning processes are explicitly represented. The results suggest that retrieval may support this integration step particularly when initial processing was insufficient, thus promoting long-term consolidation. At present, the model does not explicitly account for such time-sensitive teaching interdependencies, nor for compensatory effects based on prior processing quality—offering a promising direction for theoretical refinement.

Taken together, the results across the three studies offer important insights into the optimization of non-interactive teaching within inquiry-based, authentic science instruction. From the perspective of the Offer–Use Model of Teaching for Students' Non-Interactive Teaching (Figure 22), they demonstrate that teaching success cannot be attributed to the teaching offer alone, but emerges from the complex interplay between offer, student-related factors, and teaching context. Each modification—drawing, distribution, and retrieval—served as a theoretically and empirically grounded variation of the teaching offer and revealed specific affordances and limitations. While the combination with drawing improved immediate conceptual knowledge and the addition of retrieval practice proved beneficial for students with low-quality generative explanations regarding lasting conceptual knowledge, the distribution of non-interactive teaching showed no overall advantage. These results highlight that the optimization of non-interactive teaching is possible, but not universally effective across all students or design formats. Instead, the findings point to the importance of teaching coherence, individual learner characteristics, and the quality of prior cognitive processing in shaping whether and how teaching offers are successfully taken up and used. In this sense, the present work provides a differentiated answer to the overarching research question by demonstrating that optimizing non-interactive teaching requires not only strong instructional design, but also a deeper understanding of how students engage with and respond to the teaching offer in complex authentic classroom settings.

Figure 22

Offer-Use Model of Teaching for Students' Non-Interactive Teaching With Results From Study 1, 2, and 3



Note. Adapted from Vieluf et al. (2020) and Fiorella & Mayer (2016).

6.1.3 Implications for the Offer–Use Model of Teaching for Students' Non-Interactive Teaching

The results across the three studies provide clear evidence that optimizing non-interactive teaching is, in principle, possible. However, they also reveal that such optimization is not equally effective across teaching contexts and for all students. While certain modifications—such as combining non-interactive teaching with drawing or retrieval—proved beneficial in specific contexts or for specific learners, others showed no overall advantage or only helped students with particular learning characteristics. These differentiated effects suggest that the same teaching offer may be taken up and used very differently, depending on the student, the design, and the surrounding teaching conditions.

These results suggest that the current Offer–Use Model of Teaching for Students' Non-Interactive Teaching (Figure 3) does not sufficiently capture potentially relevant influencing factors and relationships specific to non-interactive teaching. For example, prior research has shown that parental involvement in school-related processes is associated with students' academic achievement and motivation (see Täschner et al., 2021), which may also influence the effectiveness of non-interactive teaching in the classroom context. Moreover, existing studies have shown that the effectiveness of non-interactive teaching may vary depending on the subject domain (Sibley et al., 2024) and the age of learners (Brod, 2021). Additionally, contextual factors at the school level may shape whether and how students can benefit from teaching activities such as non-interactive teaching. For example, in schools in socially deprived areas, students may face language-related challenges—particularly if many of the students have a migration background (Klein, 2017). These challenges may already impair early phases of learning, such as the selection and encoding of information in sensory memory, and may subsequently affect all downstream processes. Such difficulties could influence the uptake and use of teaching offers, students' motivation, the co-construction process including non-interactive teaching, and ultimately, students' learning outcomes.

Following these considerations, the present Offer–Use Model of Teaching for Students' Non-Interactive Teaching (Figure 3) is expanded to include an outer contextual framework—encompassing, for example, the students' direct social environment and the school context (see Figure 23). This extension is in line with the broader conception of teaching proposed by Vieluf et al. (2020), which emphasizes that teaching interactions are embedded in and shaped by multiple layers of systemic, institutional, and social conditions.

These additional contextual layers do not alter the internal logic of the original model, but they expand its explanatory scope. By explicitly integrating factors such as the school environment, family-related influences, and broader social or structural conditions, the model becomes better equipped to account for the variability observed in how teaching offers are taken up and used. In light of the present findings, such contextual framing is particularly relevant: whether a given teaching offer—including non-interactive teaching—is effective may depend not only on its design and implementation, but also on the conditions under which it is received and enacted in real classroom settings. The adapted model thus allows for a more ecologically valid representation of teaching processes and helps explain why the same learning activity may lead to different outcomes across schools, classrooms, or student groups.

It is therefore essential not to view the generative learning activity of non-interactive teaching—situated within the teaching offer—in isolation, but rather as one element within a dynamic and adaptive system. All levels and their respective components interact with one another. If certain factors within the model change, the learning activity of non-interactive teaching may need to be adjusted accordingly—and vice versa. For example, if tablets are no longer available at a given school, this may require adapting the implementation of non-interactive teaching (see Studies 1, 2, and 3 in this dissertation), which in turn could affect students' actions, their motivation, their learning processes, and ultimately, their learning outcomes.

However, the Expanded Offer–Use Model of Teaching for Students' Non-Interactive Teaching (Figure 23) reaches its limits when it comes to representing temporal and processual dynamics. If one were to depict varying emphases, interactions, and influencing factors—depending on their relative dominance and intensity—through proportionally scaled elements, the model would need to be visualized multiple times to reflect, for instance, different stages of the learning process. Ideally, it would take the form of a dynamic, video-like representation. Such a dynamic representation would allow for modeling, for example, how teaching offers unfold across multiple phases of a learning sequence, how their effects vary depending on previous engagement, and how students' use of the offers may change over time. Moreover, a dynamic model would allow for tracing how immediate learning can evolve into lasting learning over time. This perspective could help account for differential uptake, process quality, and learner-internal variability—factors that are central to understanding why the same learning activity may lead to diverse outcomes across individuals and contexts.

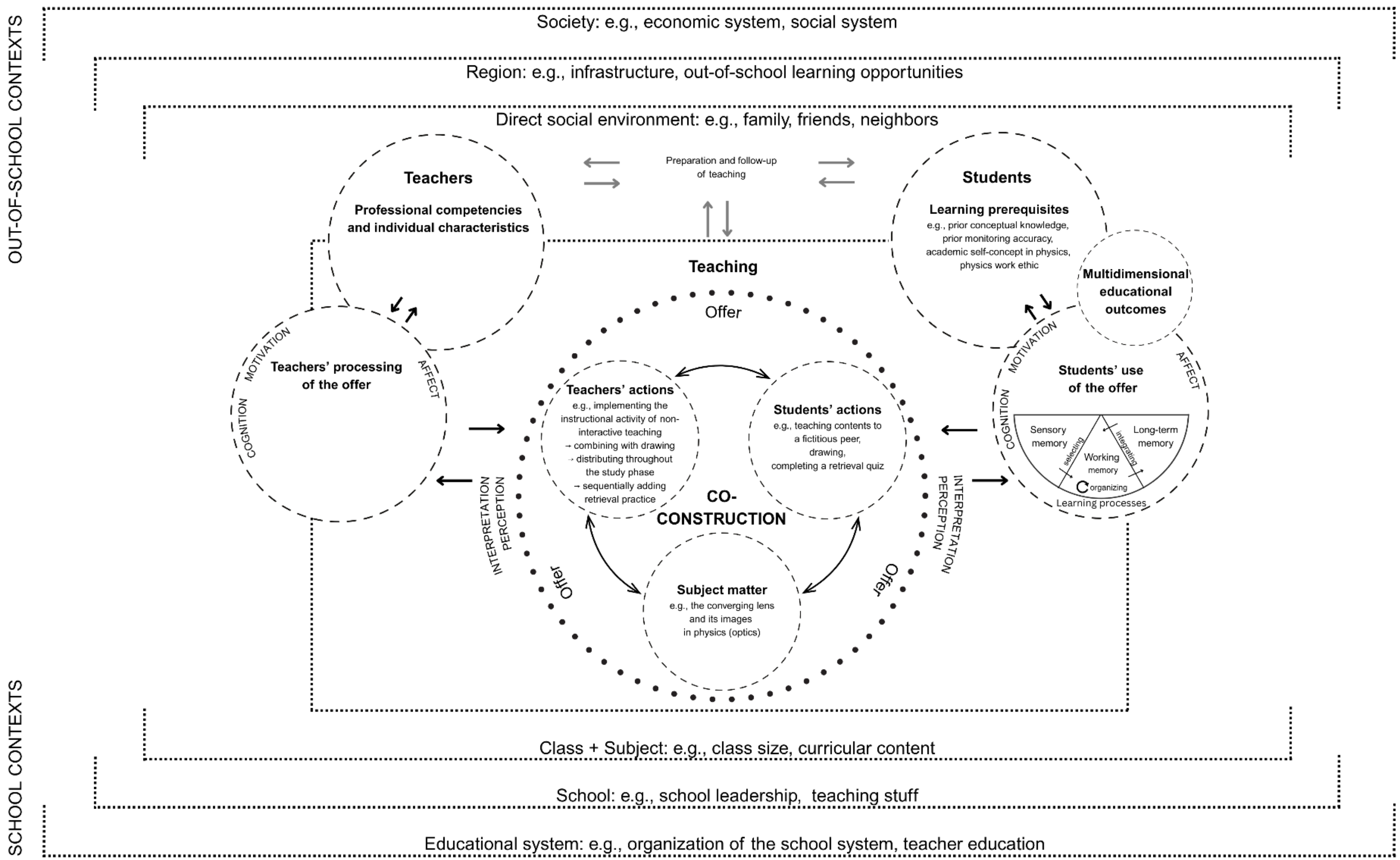
As with any model, the Expanded Offer-Use Model of Teaching for Students' Non-Interactive Teaching (Figure 23)—whether static or dynamic—inevitably entails a reduction of teaching complexity and captures only selected aspects of teaching, learning, and contextual factors. Looking ahead, it may be worthwhile to consider how the concept of teaching quality could be more obviously integrated into future iterations of the model, as previous research has demonstrated that teaching quality is a central determinant of student learning outcomes (e.g., Alp Christ et al., 2022; see also Holzberger & Schiepe-Tiska, 2021).

Within the current model (see Figure 23), the concept of learning has so far been grounded in the tradition of teaching and learning research, focusing primarily on students' performance (e.g., as in the SOI model of generative learning; Fiorella & Mayer, 2016).

However, in light of the Expanded Offer–Use Model of Teaching for Students' Non-Interactive Teaching (Figure 23), this narrow understanding appears no longer sufficient. Once broader contextual factors and processes are considered, it becomes necessary to adopt a more holistic and pedagogically grounded conception of learning—one that does not reduce learning to performance outcomes alone (see also Praetorius & Gräsel, 2021). In line with Vieluf et al. (2020), it is more appropriate to speak of multidimensional educational outcomes (see Figure 23), which include but are not limited to students' academic performance. This broader understanding is particularly relevant for modeling and evaluating non-interactive teaching as a generative activity within real-world classroom settings. From a pedagogical perspective, learning also involves engaging with the contents in a way that shapes how students relate to themselves and the world (Göhlich et al., 2014). It includes personal development, the formation of values and attitudes, critical reflection, and the ability to take informed action. Moreover, it encompasses interest in the subject matter, the development of learning strategies (learning how to learn), and the ability to engage meaningfully with societal questions and challenges. In this sense, learning is not only about what students know, but also about who they become in the process of acquiring and applying that knowledge (Göhlich et al., 2014; Lazarides & Raufelder, 2025). Ultimately, such a broadened perspective on learning aligns with the educational and developmental mission of schools, which are not solely tasked with fostering academic achievement, but also with supporting the holistic growth and maturation of the students. Understanding non-interactive teaching in this broader educational sense may thus help clarify its role not only in fostering conceptual knowledge, but also in contributing to the wider goals of education.

Figure 23

Expanded Offer-Use Model of Teaching for Students' Non-Interactive Teaching



Note. Adapted from Vieluf et al. (2020) and Fiorella & Mayer (2016).

6.2 Implications for Educational Research

Building on the results discussed above and their theoretical embedding within the Expanded Offer–Use Model of Teaching for Students' Non-Interactive Teaching (Figure 23), the present section reflects on the broader implications of this dissertation for educational research. Specifically, it addresses how this dissertation could contribute to future research agendas, offer methodological perspectives on the study of teaching and learning in authentic classroom contexts, and advance conceptual developments in the field.

6.2.1 Relevance and Placement Within the Field

Research on non-interactive teaching is still frequently conducted in laboratory settings with university students (e.g., Fiorella & Mayer, 2013, 2014; Hoogerheide et al., 2014), or—less commonly—in classroom laboratories with school students (e.g., Jacob et al., 2022) or in school contexts without curricular integration (e.g., Hoogerheide, Visee, et al., 2019). While these approaches offer advantages in terms of experimental control and internal validity, they raise questions about ecological validity and generalizability when it comes to learning processes in real school contexts.

If the goal is to draw meaningful conclusions about students' learning in schools, then research should ideally be conducted within the authentic settings in which that learning occurs. Compared to controlled laboratory conditions, classrooms typically involve more heterogeneous student populations, more complex social dynamics, and a range of situational influences that cannot easily be simulated (e.g., Dinsmore & Alexander, 2012; see also Figure 23). These contextual factors may not only affect how learning activities are implemented, but also whether and how students engage with them—ultimately influencing the outcomes observed.

Conducting classroom-based research undoubtedly requires considerable effort: it involves—for example—securing various permissions and consents, coordinating with school leadership and teachers, aligning with school schedules, preparing curricular materials, conducting related lessons and research, and managing logistical challenges. Nonetheless, in order to generate findings that are both robust and relevant for practice, this effort is essential. In this respect, educational research and educational practice must work together more closely to ensure that research reflects the complexity of teaching and learning in authentic school settings.

Against this backdrop, the present dissertation responds to the need for ecologically valid research by systematically investigating non-interactive teaching in authentic classroom settings. All three studies were embedded in regular science lessons at secondary schools and thus contribute novel, context-sensitive findings to the field.

6.2.2 Methodological Reflections on Research in Authentic Classroom Settings

The present dissertation contributes to the body of research that aims to investigate teaching processes within authentic school settings. Compared to laboratory studies, research in real classrooms allows for more ecologically valid insights into how learning activities unfold under typical conditions. However, it also poses substantial methodological challenges—particularly regarding the extent to which research designs interfere with the natural dynamics of teaching.

This tension became particularly evident in Study 2 of this dissertation following a 3 × 2 between-participants design (learning activity × timing), where two experimental groups within one class followed different timelines during the lesson, separated by partitions in the same classroom. Moreover, for the learning activity the students of these experimental groups were randomized to one of three learning activities, which were also conducted in the same classroom. While this design allowed for controlled comparisons, it also represented a considerable structural intrusion into the teaching flow. Such intrusions may possibly affect, for example, students' sense of the coherence of teaching, social interactions, peer dynamics, or motivation which—compared to fully authentic classroom settings—could lead to potentially biased data on learning outcomes. Consequently, researchers must carefully consider the potential impact of their designs on both the teaching process and the interpretability of the data.

These methodological reflections highlight the importance of balancing experimental control with pedagogical realism. In classroom-based research, implementation must remain feasible, acceptable, and meaningful within the everyday routines of school life. Rather than viewing teaching authenticity and methodological rigor as mutually exclusive, future research should seek to integrate both perspectives through thoughtful study design. This includes planning for minimal disruption, working closely with teachers, and aligning interventions with curricular goals and existing classroom structures. Ultimately, advancing research in authentic school contexts requires a methodological mindset that values contextual sensitivity as much as internal validity.

6.2.3 Deliberate Research Focus and Analytical Scope

In the present dissertation, the research focus was deliberately narrowed to a well-defined learning activity—non-interactive teaching—and the outcome variables students' conceptual knowledge and monitoring accuracy, regarding both immediate and lasting learning. This choice follows a well-established tradition in teaching and learning research, which prioritizes clearly delineated experimental conditions and outcome measures in order to enable precise analyses of specific instructional mechanisms. Focusing on a small segment of the learning process allowed for a deeper understanding of how students engage with non-interactive teaching and how this activity can be optimized through modifications—drawing, distribution, or retrieval—for immediate and lasting effects.

At the same time, this focused perspective raises important questions about what is included in the research focus and what is left out. As the Expanded Offer–Use Model of

Teaching for Students' Non-Interactive Teaching (Figure 23) illustrates, classroom teaching is embedded in a complex network of interacting components—including not only students and subject matter, but also teachers, school-level conditions, and the broader learning environment. Accordingly, selecting a narrow research focus also means defining the analytical boundaries of the investigation, specifying which parts of the teaching interaction are foregrounded and which components are purposefully held constant or left unmeasured. In light of the complexity of authentic classroom settings, such decisions should be made with care: depending on the research question, it may be necessary to broaden the scope and include additional model components to capture potential influencing factors more fully. Integrating such elements—for example, teachers' professional competencies and individual characteristics, teaching quality, or school context variables—may help to build a more comprehensive picture of how and under which conditions specific instructional activities like non-interactive teaching unfold and take effect in real classrooms (see also Vieluf et al., 2020).

6.3 Implications for Educational Practice

In addition to its contributions to educational research, this dissertation also yields implications for educational practice—particularly in the domains of classroom teaching, school development, and teacher education. These implications are elaborated in the following subsections, with the specific focus on the generative learning activity of non-interactive teaching.

6.3.1 Implications for Classroom Teaching

The studies of this dissertation demonstrated that non-interactive teaching combined with drawing is an effective learning strategy and that its combination with retrieval practice holds particular promise for supporting lasting learning. Importantly, these activities are low-threshold in terms of resources and can be integrated flexibly into everyday teaching without requiring major structural changes. For instance, non-interactive teaching and drawing can be placed at the end of inquiry-based teaching units to reinforce students' conceptual understanding. Teachers may concentrate on a limited set of clearly defined key learning objectives and integrate hands-on experiments or concrete materials during their lessons. Building on this phase, teachers could present a chat message from a fictitious peer to elicit students' verbal explanations and corresponding drawings—in class or as a homework assignment (Hoogerheide, Visee, et al., 2019). To enhance long-term retention, a retrieval task—such as a follow-up quiz—may be incorporated as a subsequent step.

Importantly, it should be noted that the learning activity of non-interactive teaching and its thematic connection to the converging lens and its images is not effective in itself but depends on the quality of its implementation (Raudenbush, 2008; see also Klieme, 2022). Effective learning is shaped not merely by surface features like teaching format or method, but especially by reaching the deeper structure of teaching, where actual student learning occurs (Bohl, 2023). In the studies of this dissertation, students' learning was enhanced by non-interactive teaching combined with drawing or retrieval practice, respectively, suggesting that these activities addressed deep structures of teaching. Teachers can access these deep structures through the core dimensions of instructional quality: cognitive activation, classroom management, and a supportive learning climate (Bohl, 2023; Praetorius et al., 2020). These dimensions are also reflected in Pietsch's (2010) hierarchical model of teaching quality, which comprises four levels: (1) ensuring a productive learning climate and pedagogical structure, (2) managing classrooms effectively and varying methods, (3) fostering motivation, active learning, and knowledge transfer, and (4) differentiating teaching and promoting competence development. The generative task of non-interactive teaching would, at minimum, correspond to the third level of this model. This underscores the high teaching demands required to successfully integrate such learning activities.

The results of this dissertation highlight the potential of generative and retrieval-based strategies to enhance everyday instruction—provided they are implemented with pedagogical precision.

6.3.2 Implications for School Development

When the use of the generative learning activity of non-interactive teaching extends beyond individual classrooms and becomes embedded in collaborative teaching structures across a school, it enters the domain of teaching development, as one part of school development alongside organizational development (i.e., developing the organization from within) and professional development (Rolff, 2018). These three dimensions are closely interlinked and should be understood as part of a systemic approach to improving schools from within (Rolff, 2018). For instance, when science teachers collaborate to implement non-interactive teaching, jointly plan lesson sequences, or co-design task formats tailored to subject-specific contents, their efforts contribute both to teaching and organizational development. Such practices inevitably influence teachers' professional routines, which connects this process to professional development, for example through reciprocal classroom observations among colleagues.

Overall, the implementation of non-interactive teaching has the potential to trigger school development processes—particularly when key conditions for success are taken into account (Haenisch & Steffens, 2017). For example, these might include a clear strategic focus on generative learning activities combined with retrieval practice, the formation of teacher teams, and systematic reflection formats such as sharing outcomes in subject or staff meetings. Such practices could help other teachers gradually engage with and adopt the approach. School leadership plays a crucial role in this process by legitimizing the initiative and facilitating structured opportunities for professional exchange. Also important is a student-centered orientation, for instance by ensuring connections to students' lifeworlds through mock-up chats and the use of tablets regarding non-interactive teaching.

In schools located in socially deprived areas, such high process quality as well as high outcome quality—for example, through successful teaching enabled by non-interactive teaching—could turn a school into an unexpectedly high-performing one. Such a school may ultimately require less development than a school situated in a more privileged socio-spatial context but showing low levels of both process and outcome quality. However, such a consideration depends on the selected focus and the specific indicators of process and outcome quality used (Bremm et al., 2016).

In this sense, the dissemination and systemic integration of generative activities potentially combined with retrieval practice, may not only foster teaching improvement but also serve as a catalyst for sustainable school development.

6.3.3 Implications for Teacher Education

For generative learning activities such as non-interactive teaching to be used effectively in classroom teaching, teachers must be both familiar with these strategies and capable of implementing them in a pedagogically meaningful way. This can be supported across all phases of teacher education. Therefore, teacher education plays a key role in ensuring that future and practicing teachers are equipped not only to apply such strategies, but to do so with pedagogical precision. It should be noted that the approaches described here represent only one specific teaching focus and are explicitly not intended to reflect the full scope or priorities of teacher education in these phases.

In the first phase of teacher education at university, prospective teachers should be introduced to the theoretical foundations and empirical evidence of generative learning strategies and retrieval practice. These learning activities should be presented within the broader context of contextual factors and teaching quality. The Expanded Offer-Use Model of Teaching for Students' Non-Interactive Teaching (see Figure 23) may serve as a helpful framework in this regard. However, it is important to ensure that prospective teachers do not interpret their role solely in terms of the offer side, and thus risk neglecting their broader teaching responsibility (see Kohler & Wacker, 2013). Moreover, to foster active understanding, prospective teachers should have the opportunity to try out these activities firsthand during their university courses and also design lesson segments in their respective subject areas that incorporate non-interactive teaching and retrieval practice.

In the second phase of teacher education (i.e., preparatory service), the focus should be on applying these strategies in authentic teaching settings. Under the guidance of mentors and seminar instructors, teacher candidates could implement non-interactive teaching, drawing, and sequential retrieval practice in real lessons and reflect on their implementation, including subject-specific considerations. Collaborative formats such as joint lesson planning or reciprocal lesson observations could help bridge the gap between theoretical knowledge, self-experienced learning activities, and their practical implementation in the classroom.

In the third phase, continuing professional development should provide opportunities for experienced teachers to become familiar with, further develop, and refine their use of generative learning strategies and retrieval practice. Since some teachers may not yet be aware of approaches such as non-interactive teaching and drawing, or their combination with retrieval practice, access to these strategies should be supported through practice-oriented formats such as workshops, teaching guides, or subject-specific handouts. Professional learning formats such as collegial exchange, guided reflection, or subject-based discussion groups can help teachers share experiences, adapt these approaches to their subject-specific contexts, and integrate them meaningfully into their everyday teaching. To ensure that such formats lead to lasting improvements in teaching quality and student learning, professional development must be designed to address multiple levels of impact—from participants' satisfaction and professional beliefs, to changes in teaching practice and, ultimately, supporting student learning (Lipowsky & Rzejak, 2021; Rzejak et al., 2020). Accordingly, teacher education should also emphasize the importance of teaching quality for student learning and integrate opportunities for teachers to reflect on the impact of their own teaching behaviors (Lipowsky & Rzejak, 2021).

Across all three phases, it is essential that teacher education does not merely introduce models of learning strategies, but actively engages teachers in critically examining, adapting, and applying them in light of their subject matter, student needs, and teaching contexts. Drawing on Lipowsky's framework of professional development (Lipowsky & Rzejak, 2021), it becomes evident that improving teaching quality must be a central goal across all phases of teacher education, as it is the key condition through which student learning can be positively affected.

6.4 Limitations and Directions for Future Research

A key limitation of this study concerns the generalizability of its findings beyond the specific experimental and teaching context. Although substantial efforts were made to approximate authentic classroom conditions—for example, by conducting the intervention in authentic lessons, aligning it with the curriculum, and having all sessions and study parts conducted by the author of this dissertation—the setting likely still differed from typical teaching practice. Everyday teaching is shaped by contextual features such as established teacher–student relationships, classroom routines, time constraints, and the need to respond flexibly to students' needs. As a result, the findings are limited in terms of their generalizability to broader, authentic educational settings. Moreover, the study's focus on a single subject, a specific topic, and a narrow age group further limits generalizability, as prior research has shown that both subject matter (Sibley et al., 2024) and developmental stage (Brod, 2021) can influence the effectiveness of learning activities. To address this limitation, future research should aim to further bridge the gap between experimental rigor and authentic classroom conditions. This could be achieved by developing less complex or intrusive designs that embed generative learning activities such as non-interactive teaching more seamlessly into regular teaching. Collaborating closely with practicing teachers—for example through professional development initiatives—would allow them to adapt these strategies to their own teaching goals, subject areas, and classroom contexts. Such co-constructed designs would not only increase ecological validity but also support the long-term integration of evidence-based learning strategies into everyday practice (see also Lipowsky & Rzejak, 2021). Finally, replication studies across different subject areas (e.g., biology, mathematics, or language learning) and age groups are needed to investigate whether, under which conditions, and for which student groups non-interactive teaching—including its possible modifications—yields immediate and lasting effects.

Another limitation relates to the study design. First, the repeated use of the same conceptual knowledge test ROC-CI (Wörner et al., 2022) at all three time points ensured consistency in measurement but may have introduced recognition effects. Although answer options were randomized, we cannot fully rule out that students remembered test items rather than demonstrating conceptual understanding. This limits the interpretability of our results. Future studies could consider developing equivalent test versions to avoid potential item recognition while assessing the same conceptual content. Second, while the ROC-CI provides a validated and content-specific measure of conceptual knowledge about the converging lens (Wörner et al., 2022), it captures only a narrow slice of what students may have learned during the inquiry-based physics lesson. The lesson structure included student experimentation and generative activities, but the assessment focused mainly on conceptual knowledge recall and did not address skills more closely aligned with inquiry learning—such as reasoning about experimental outcomes or evaluating scientific evidence (see Pedaste et al., 2015). Furthermore, while drawing was a central instructional component in some conditions, no specific drawing test was used to assess students' ability to translate knowledge into visual representations, as implemented in recent research on generative learning (Fiorella, 2023a). To improve the methodological coherence of future designs, assessments should be selected or developed to reflect the broader set of intended learning goals addressed during teaching. This could include differentiated measures targeting conceptual understanding in optics, drawing-based representational skills, and practices related to inquiry learning.

Another limitation concerns the rather small effect sizes observed in the studies of this dissertation. One potential explanation for the modest effects lies in the short duration of the intervention within an authentic classroom setting, which also may not have been sufficient to foster lasting learning effects over an 8-week period. Although the intervention was brief, it still provided more time for engagement with generative tasks than most previous studies in the field (e.g., Hoogerheide et al., 2016; Lachner et al., 2022). When interpreting these results, it is important to consider both the brevity of the intervention and the typically small to medium effect sizes reported in meta-analyses on generative learning (e.g., Lachner et al., 2022; Ribosa & Duran, 2022). Against this background, the observed effects appear meaningful. Moreover, Kraft (2020) argues that even small effect sizes can be substantively important if the intervention is scalable, cost-effective, and capable of producing cumulative effects when applied at scale. Further, when compared to annual learning gains in standardized assessments in middle school—typically ranging between $d = 0.20$ and $d = 0.30$ (Bloom et al., 2008; Hill et al., 2008)—the effects observed in the studies of this dissertation appear even more substantial. A promising avenue for future research lies in examining the scalability and long-term feasibility of embedding generative and retrieval-based activities into regular classroom teaching.

6.5 Conclusion

This dissertation addressed the overarching question of how non-interactive teaching can be optimized to enhance students' conceptual understanding and monitoring accuracy in inquiry-based authentic science lessons—both in terms of immediate and lasting learning. To this end, three experimental classroom studies investigated distinct modifications of non-interactive teaching: combining it with drawing (Study 1), distributing it across the study phase (Study 2), and sequentially adding retrieval practice (Study 3). Each study targeted one theoretically and empirically grounded enhancement and contributed unique empirical insights into optimizing non-interactive teaching in real-world classroom settings.

Taken together, the findings demonstrate that non-interactive teaching can be systematically enhanced. While drawing increased students' task-specific interest and improved the quality of their explanations, thereby boosting immediate learning, distributing the activity did not yield additional benefits. Moreover, retrieval practice supported lasting learning particularly when the quality of prior generative processing was low, underscoring the importance of tailoring instructional interventions to students' needs.

These results support an expanded understanding of non-interactive teaching as part of a teaching offer that can foster students' generative learning processes also in complex school contexts. In this sense, non-interactive teaching can contribute to strengthening successful teaching and thus teaching quality. From this perspective, not only immediate performance but also lasting learning emerges as a key educational goal (OECD, 2016). The present findings suggest promising avenues for promoting lasting learning by combining generative tasks with retrieval practice. A broader teaching perspective further underscores that, beyond knowledge acquisition, generative learning activities may contribute to multidimensional learning outcomes—such as increased interest in subject matter, the development of self-regulated learning strategies, and the ability to engage meaningfully with complex societal questions and challenges (see also Klieme, 2022; Vieluf et al., 2020).

Looking ahead, this resonates with the vision laid out in the OECD Learning Compass 2030 (OECD, 2019), which emphasizes the importance of empowering students to create new value, navigate ambiguity, and act responsibly in an increasingly complex world. As a generative learning activity, non-interactive teaching—potentially enhanced through drawing and retrieval practice—could support this vision by fostering deep understanding, critical reflection, and the construction of meaningful lasting knowledge. Ensuring that such strategies become accessible and scalable in everyday classroom teaching could contribute to preparing all students to navigate the complexities of a rapidly changing world.

7

REFERENCES

7 References

Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research*, *87*(3), 659–701. https://doi.org/10.3102/0034654316689306

Aguinis, H. (1995). Statistical power problems with moderated multiple regression in management research. *Journal of Management*, *21*(6), 1141–1158. https://doi.org/10.1016/0149-2063(95)90026-8

Ainsworth, S., & Loizou, A. T. (2003). The effects of self-explaining when learning with text or diagrams. *Cognitive Science*, *27*(4), 669–681. https://doi.org/10.1207/s15516709cog2704_5

Ainsworth, S., & Scheiter, K. (2021). Learning by drawing visual representations: Potential, purposes, and practical implications. *Current Directions in Psychological Science*, *30*(1), 61–67. https://doi.org/10.1177/0963721420979582

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

Alp Christ, A., Capon-Sieber, V., Grob, U., & Praetorius, A.-K. (2022). Learning processes and their mediating role between teaching quality and student achievement: A systematic review. *Studies in Educational Evaluation*, *75*, 101209. https://doi.org/10.1016/j.stueduc.2022.101209

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist*, *73*(1), 3–25. https://doi.org/10.1037/amp0000191

Baars, M., van Gog, T., de Bruin, A., & Paas, F. (2017). Effects of problem solving after worked example study on secondary school children’s monitoring accuracy. *Educational Psychology*, *37*(7), 810–834. https://doi.org/10.1080/01443410.2016.1150419

Bareis, A., Spengler, M., Rieger, S., Gladstone, J., Seung Yang, J., Nagengast, B., Trautwein, U., & Wigfield, A. (2024). Examining the Conscientiousness × Interest Compensation (CONIC) Model with similar constructs in high school and college students. *Journal of Research in Personality*, 104462. https://doi.org/10.1016/j.jrp.2024.104462

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin*, *128*(4), 612–637. https://doi.org/10.1037/0033-2909.128.4.612

Baumert, J., Grabbe, Y., Kunter, M., Dubberke, T., Elsner, J., Furtak, E., Klusmann, U., Richter, D., & Tsai, Y.-M. (2012). *COACTIV-R Follow-up Schülerfragebogen: Dokumentation der Erhebungsinstrumente [COACTIV-R Follow-up student questionnaire: Documentation of survey instruments]*. Max-Planck-Institut für Bildungsforschung.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klusmann, U., Krauss, S., Neubrand, M., & Tsai, Y.-M. (2010). Teachers’ mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, *47*(1), 133–180. https://doi.org/10.3102/0002831209345157

Berliner, D. C. (2005). The near impossibility of testing for teacher quality. *Journal of Teacher Education*, *56*(3), 205–213. https://doi.org/10.1177/0022487105275904

Betella, A., & Verschure, P. F. M. J. (2016). The affective slider: A digital self-assessment scale for the measurement of human emotions. *PLOS ONE*, *11*(2), e0148037. https://doi.org/10.1371/journal.pone.0148037

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

Bjork, R. A., & Bjork, E. L. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). Worth Publishers.

Bloom, H. S., Hill, C. J., Black, A. R., & Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Journal of Research on Educational Effectiveness*, *1*(4), 289–328. https://doi.org/10.1080/19345740802400072

Bohl, T. (2023). Umgang mit Heterogenität im Unterricht: Forschungsbefunde und didaktische Implikationen [Dealing with heterogeneity in the classroom: Research findings and didactic implications]. In T. Bohl, J. Budde, & M. Rieger-Ladich (Eds.), *Umgang mit Heterogenität in Schule und Unterricht: Grundlagentheoretische Beiträge, empirische Befunde und didaktische Reflexionen* (2. aktualisierte Auflage, pp. 263–279). UTB. https://doi.org/10.36198/9783838559667

Boshuizen, H. P. A., & Schmidt, H. G. (1992). On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices. *Cognitive Science*, *16*(2), 153–184. https://doi.org/10.1207/s15516709cog1602_1

Bremm, N., Klein, E. D., & Racherbäumer, K. (2016). Schulen in „schwieriger“ Lage?! Begriffe, Forschungsbefunde und Perspektiven [Schools in “challenging” contexts?! Terms, research findings and perspectives]. *DDS – Die Deutsche Schule*, *108*(4), 323–339.

Brod, G. (2021). Generative learning: Which strategies for what age? *Educational Psychology Review*, *33*(4), 1295–1318. https://doi.org/10.1007/s10648-020-09571-9

Brod, G. (2024). There are multiple paths to personalized education, and they should be combined. *Current Directions in Psychological Science*, *33*(3), 153–158. https://doi.org/10.1177/09637214241242459

Cadima, J., Leal, T., & Burchinal, M. (2010). The quality of teacher–student interactions: Associations with first graders' academic and behavioral outcomes. *Journal of School Psychology*, *48*(6), 457–482. https://doi.org/10.1016/j.jsp.2010.09.001

Callender, A. A., Franco-Watkins, A. M., & Roberts, A. S. (2016). Improving metacognition in the classroom through instruction, training, and feedback. *Metacognition and Learning*, *11*(2), 215–235. https://doi.org/10.1007/s11409-015-9142-6

Carpenter, S. K., Cepeda, N. J., Rohrer, D., Kang, S. H. K., & Pashler, H. (2012a). Using spacing to enhance diverse forms of learning: Review of recent research and implications for instruction. *Educational Psychology Review*, *24*(3), 369–378. https://doi.org/10.1007/s10648-012-9205-z

Carpenter, S. K., Cepeda, N. J., Rohrer, D., Kang, S. H. K., & Pashler, H. (2012b). Using spacing to enhance diverse forms of learning: Review of recent research and implications for instruction. *Educational Psychology Review*, *24*(3), 369–378. https://doi.org/10.1007/s10648-012-9205-z

Castro-Alonso, J. C., de Koning, B. B., Fiorella, L., & Paas, F. (2021). Five strategies for optimizing instructional materials: Instructor- and learner-managed cognitive load. *Educational Psychology Review*, *33*(4), 1379–1407. https://doi.org/10.1007/s10648-021-09606-9

Chi, M. T. H., & Roscoe, R. D. (2002). The processes and challenges of conceptual change. In M. Limón & L. Mason (Eds.), *Reconsidering conceptual change: Issues in theory and practice* (pp. 3–27). Springer Netherlands. https://doi.org/10.1007/0-306-47637-1_1

Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, *49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

Chua, E. F., Schacter, D. L., & Sperling, R. A. (2009). Neural correlates of metamemory: A comparison of feeling-of-knowing and retrospective confidence judgments. *Journal of Cognitive Neuroscience*, *21*(9), 1751–1765. https://doi.org/10.1162/jocn.2009.21123

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Academic press.

Cooper, M. M., Stieff, M., & DeSutter, D. (2017). Sketching the invisible to predict the visible: From drawing to modeling in chemistry. *Topics in Cognitive Science*, *9*(4), 902–920. https://doi.org/10.1111/tops.12285

Cox, R. (1999). Representation construction, externalised cognition and individual differences. *Learning and Instruction*, *9*(4), 343–363. https://doi.org/10.1016/S0959-4752(98)00051-6

Cromley, J. G. (2020). Learning from multiple representations: Roles of task interventions and individual differences. In P. Van Meter, A. List, D. Lombardi, & P. Kendeou (Eds.), *Handbook of learning from multiple representations and perspectives* (1st ed., pp. 62–75). Routledge. https://doi.org/10.4324/9780429443961

Cromley, J. G., Du, Y., & Dane, A. P. (2020). Drawing-to-learn: Does meta-analysis show differences between technology-based drawing and paper-and-pencil drawing? *Journal of Science Education and Technology*, *29*(2), 216–229. https://doi.org/10.1007/s10956-019-09807-6

Cromley, J. G., Snyder-Hogan, L. E., & Luciw-Dubas, U. A. (2010). Cognitive activities in complex science text and diagrams. *Contemporary Educational Psychology*, *35*(1), 59–74. https://doi.org/10.1016/j.cedpsych.2009.10.002

Cuddy, L. J., & Jacoby, L. L. (1982). When forgetting helps memory: An analysis of repetition effects. *Journal of Verbal Learning and Verbal Behavior*, *21*(4), 451–467. https://doi.org/10.1016/S0022-5371(82)90727-7

De Fruyt, F., Van Leeuwen, K., De Bolle, M., & De Clercq, B. (2008). Sex differences in school performance as a function of conscientiousness, imagination and the mediating role of problem behaviour. *European Journal of Personality*, *22*(3), 167–184. https://doi.org/10.1002/per.675

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., Chinn, C. A., Fischer, F., Gobert, J., Hmelo-Silver, C. E., Koedinger, K. R., Krajcik, J. S., Kyza, E. A., Linn, M. C., Pedaste, M., Scheiter, K., & Zacharia, Z. C. (2023). Let’s talk evidence – The case for combining inquiry-based and direct instruction. *Educational Research Review*, *39*, 100536. https://doi.org/10.1016/j.edurev.2023.100536

Decristan, J., Kunter, M., Fauth, B., Büttner, G., Hardy, I., & Hertel, S. (2016). What role does instructional quality play for elementary school children’s science competence? A focus on students at risk. *Journal for Educational Research Online*, *8*(1), 66–89. https://doi.org/10.25656/01:12032

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*(1), 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Dinsmore, D. L., & Alexander, P. A. (2012). A critical discussion of deep and surface processing: What it means, how it is measured, the role of context, and model specification. *Educational Psychology Review*, *24*(4), 499–567. https://doi.org/10.1007/s10648-012-9198-7

Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction*, *22*(4), 271–280. https://doi.org/10.1016/j.learninstruc.2011.08.003

Ebersbach, M., Lachner, A., Scheiter, K., & Richter, T. (2022). Using spacing to promote lasting learning in educational contexts. *Zeitschrift Für Entwicklungspsychologie Und Pädagogische Psychologie*, *54*(4), 151–163. https://doi.org/10.1026/0049-8637/a000259

Enders, C. K. (2022). *Applied missing data analysis*. Guilford Publications.

Endres, T., Carpenter, S., & Renkl, A. (2024). Constructive retrieval: Benefits for learning, motivation, and metacognitive monitoring. *Learning and Instruction*, *94*, 101974. https://doi.org/10.1016/j.learninstruc.2024.101974

Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Büttner, G. (2014). Student ratings of teaching quality in primary school: Dimensions and prediction of student outcomes. *Learning and Instruction*, *29*, 1–9. https://doi.org/10.1016/j.learninstruc.2013.07.001

Fend, H. (Ed.). (2008). *Schule gestalten: Systemsteuerung, Schulentwicklung und Unterrichtsqualität [Shaping schools: System governance, school development and teaching quality]*. VS Verlag. https://doi.org/10.1007/978-3-531-90867-0

Finn, B. (2015). Measuring motivation in low-stakes assessments. *ETS Research Report Series*, *2015*(2), 1–17. https://doi.org/10.1002/ets2.12067

Fiorella, L. (2022). Learning by explaining after pauses in video lectures: Are provided visuals a scaffold or a crutch? *Applied Cognitive Psychology*, *36*(5), 1142–1149. https://doi.org/10.1002/acp.3994

Fiorella, L. (2023a). Fostering knowledge building in learning by teaching: A test of the drawing-facilitates-explaining hypothesis. *Applied Cognitive Psychology*, *37*(5), 1124–1138. https://doi.org/10.1002/acp.4100

Fiorella, L. (2023b). Making sense of generative learning. *Educational Psychology Review*, *35*, 50. https://doi.org/10.1007/s10648-023-09769-7

Fiorella, L., & Jaeger, A. J. (2023). Are there metacognitive benefits of learner- and instructor-generated visualizations? *Applied Cognitive Psychology*, *37*(6), 1430–1443. https://doi.org/10.1002/acp.4135

Fiorella, L., Jaeger, A. J., Capobianco, A., & Burnett, A. (2024). "My drawing is quite different!" Drawbacks of comparing generative drawings to instructional visuals. *Contemporary Educational Psychology*, *77*, 102277. https://doi.org/10.1016/j.cedpsych.2024.102277

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

Fiorella, L., & Mayer, R. E. (2013). The relative benefits of learning by teaching and teaching expectancy. *Contemporary Educational Psychology*, *38*(4), 281–288. https://doi.org/10.1016/j.cedpsych.2013.06.001

Fiorella, L., & Mayer, R. E. (2014). Role of expectations and explanations in learning by teaching. *Contemporary Educational Psychology*, *39*(2), 75–85. https://doi.org/10.1016/j.cedpsych.2014.01.001

Fiorella, L., & Mayer, R. E. (2015). *Learning as a generative activity*. Cambridge university press.

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

Fiorella, L., & Zhang, Q. (2018). Drawing boundary conditions for learning by drawing. *Educational Psychology Review*, *30*(3), 1115–1137. https://doi.org/10.1007/s10648-018-9444-8

Flegr, S., Kuhn, J., & Scheiter, K. (2023). When the whole is greater than the sum of its parts: Combining real and virtual experiments in science education. *Computers & Education*, *197*, 104745. https://doi.org/10.1016/j.compedu.2023.104745

Fleming, S. M., Massoni, S., Gajdos, T., & Vergnaud, J.-C. (2016). Metacognition about the past and future: Quantifying common and distinct influences on prospective and retrospective judgments of self-performance. *Neuroscience of Consciousness*, *2016*(1), niw018. https://doi.org/10.1093/nc/niw018

Fukaya, T. (2013). Explanation generation, not explanation expectancy, improves metacomprehension accuracy. *Metacognition and Learning*, *8*(1), 1–18. https://doi.org/10.1007/s11409-012-9093-0

Furr, R. M., & Rosenthal, R. (2003). Evaluating theories efficiently: The nuts and bolts of contrast analysis. *Understanding Statistics*, *2*(1), 33–67. https://doi.org/10.1207/S15328031US0201_03

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

Göhlich, M., Wulf, C., & Zirfas, J. (2014). Pädagogische Zugänge zum Lernen [Pedagogical approaches to learning]. In M. Göhlich, C. Wulf, & J. Zirfas (Eds.), *Pädagogische Theorien des Lernens* (2. Aufl, pp. 7–19). Beltz Juventa.

Goldwater, M. B., & Schalk, L. (2016). Relational categories as a bridge between cognitive and educational research. *Psychological Bulletin*, *142*(7), 729–757. https://doi.org/10.1037/bul0000043

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, *60*(1), 549–576. https://doi.org/10.1146/annurev.psych.58.110405.085530

Greenland, S. (2000). Principles of multilevel modelling. *International Journal of Epidemiology*, *29*(1), 158–167. https://doi.org/10.1093/ije/29.1.158

Gudjons, H., & Traub, S. (Eds.). (2020). *Pädagogisches Grundwissen [Basic knowledge in pedagogy]* (13. aktual. Aufl.). Verlag Julius Klinkhardt. https://doi.org/10.36198/9783838555232

Guerrero, T. A., & Wiley, J. (2021). Expecting to teach affects learning during study of expository texts. *Journal of Educational Psychology*, *113*(7), 1281–1303. https://doi.org/10.1037/edu0000657

Gutierrez, A. P., Schraw, G., Kuch, F., & Richmond, A. S. (2016). A two-process model of metacognitive monitoring: Evidence for general accuracy and error factors. *Learning and Instruction*, *44*, 1–10. https://doi.org/10.1016/j.learninstruc.2016.02.006

Gutierrez de Blume, A. P., Schraw, G., Kuch, F., Richmond, A. S., Gutierrez de Blume, A. P., Schraw, G., Kuch, F., & Richmond, A. S. (2021). General accuracy and general error factors in metacognitive monitoring and the role of time on task in predicting metacognitive judgments. *CES Psicología*, *14*(2), 179–208. https://doi.org/10.21615/cesp.5494

Haenisch, H., & Steffens, U. (2017). Schlüsselfaktoren für die Entwicklung von Schulen [Key factors for school development]. In U. Steffens, K. M. Merki, & H. Fend (Eds.), *Schulgestaltung: Aktuelle Befunde und Perspektiven der Schulqualitäts- und Schulentwicklungsforschung* (pp. 160–184). Waxmann Verlag.

Hall, V. C., Bailey, J., & Tillman, C. (1997). Can student-generated illustrations be worth ten thousand words? *Journal of Educational Psychology*, *89*(4), 677–681. https://doi.org/10.1037/0022-0663.89.4.677

Händel, M., Harder, B., & Dresel, M. (2020). Enhanced monitoring accuracy and test performance: Incremental effects of judgment training over and above repeated testing. *Learning and Instruction*, *65*, 101245. https://doi.org/10.1016/j.learninstruc.2019.101245

Hayes, A. F., & Montoya, A. K. (2017). A tutorial on testing, visualizing, and probing an interaction involving a multicategorical variable in linear regression analysis. *Communication Methods and Measures*, *11*(1), 1–30. https://doi.org/10.1080/19312458.2016.1271116

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives*, *2*(3), 172–177. https://doi.org/10.1111/j.1750-8606.2008.00061.x

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

Holzberger, D., & Schiepe-Tiska, A. (2021). Is the school context associated with instructional quality? The effects of social composition, leadership, teacher collaboration, and school climate. *School Effectiveness and School Improvement*, *32*(3), 465–485. https://doi.org/10.1080/09243453.2021.1913190

Hoogerheide, V., Deijkers, L., Loyens, S. M. M., Heijltjes, A., & Van Gog, T. (2016). Gaining from explaining: Learning improves from explaining to fictitious others on video, not from writing to them. *Contemporary Educational Psychology*, *44–45*, 95–106. https://doi.org/10.1016/j.cedpsych.2016.02.005

Hoogerheide, V., Loyens, S. M., & Van Gog, T. (2014). Effects of creating video-based modeling examples on learning and transfer. *Learning and Instruction*, *33*, 108–119. https://doi.org/10.1016/j.learninstruc.2014.04.005

Hoogerheide, V., Renkl, A., Fiorella, L., Paas, F., & van Gog, T. (2019). Enhancing example-based learning: Teaching on video increases arousal and improves problem-solving performance. *Journal of Educational Psychology*, *111*(1), 45–56. https://doi.org/10.1037/edu0000272

Hoogerheide, V., Visee, J., Lachner, A., & van Gog, T. (2019). Generating an instructional video as homework activity is both effective and enjoyable. *Learning and Instruction*, *64*, 101226. https://doi.org/10.1016/j.learninstruc.2019.101226

Ifenthaler, D., & Seel, N. M. (2011). A longitudinal perspective on inductive reasoning tasks. Illuminating the probability of change. *Learning and Instruction*, *21*(4), 538–549. https://doi.org/10.1016/j.learninstruc.2010.08.004

Jacob, L., Lachner, A., & Scheiter, K. (2020). Learning by explaining orally or in written form? Text complexity matters. *Learning and Instruction*, *68*, 101344. https://doi.org/10.1016/j.learninstruc.2020.101344

Jacob, L., Lachner, A., & Scheiter, K. (2021). Does increasing social presence enhance the effectiveness of writing explanations? *PloS One*, *16*(4), e0250406. https://doi.org/10.1371/journal.pone.0250406

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

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative Big-Five trait taxonomy: History, measurement, and conceptual issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (pp. 114–158). Guilford Press.

John, O. P., & Srivastava, S. (1999). The big five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (Subsequent edition, pp. 102–138). Guilford Publications.

Johnson-Laird, P. N. (1989). Mental models. In M. I. Posner (Ed.), *Foundations of cognitive science* (pp. 469–499). The MIT Press.

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review*, *19*(4), 509–539. https://doi.org/10.1007/s10648-007-9054-3

Kalyuga, S. (2014). The expertise reversal principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 576–597). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.028

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

Karpicke, J. D. (2017). Retrieval-based learning: A decade of progress. In J. H. Byrne (Ed.), *Learning and memory: A comprehensive reference* (2nd ed., pp. 487–514). Academic Press. https://doi.org/10.1016/B978-0-12-809324-5.21055-9

Keller, S. D., Markwalder, U., Endres, T., & Praetorius, A.-K. (2025). Die Unterstützung des Konsolidierens – eine zentrale, aber vernachlässigte Dimension von Unterrichtsqualität [Supporting consolidation – A central but neglected dimension of teaching quality]. *Unterrichtswissenschaft*. https://doi.org/10.1007/s42010-025-00229-5

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. https://doi.org/10.1207/s15326985ep4102_1

Klafki, W. (1995). Didactic analysis as the core of preparation of instruction [Didaktische Analyse als Kern der Unterrichtsvorbereitung]. *Journal of Curriculum Studies*, *27*(1), 13–30. https://doi.org/10.1080/0022027950270103

Klein, E. D. (2017). Bedingungen und Formen erfolgreicher Schulentwicklung in Schulen in sozial deprivierter Lage. Eine Expertise im Auftrag der Wübben Stiftung [Conditions and forms of successful school development in schools in socially deprived contexts. An expert report commissioned by the Wübben Foundation]. *SHIP Working Paper Reihe*, *1*. https://doi.org/10.17185/DUEPUBLICO/44384

Klepsch, M., & Seufert, T. (2021). Making an effort versus experiencing load. *Frontiers in Education*, *6*, 645284. https://doi.org/10.3389/feduc.2021.645284

Klieme, E. (2022). Unterrichtsqualität [Teaching quality]. In M. Harring, C. Rohlfs, & M. Gläser-Zikuda (Eds.), *Handbuch Schulpädagogik* (2., aktualisierte underweiterte Auflage, pp. 411–575). Waxmann.

Kobayashi, K. (2024). Interactive learning effects of preparing to teach and teaching: A meta-analytic approach. *Educational Psychology Review*, *36*(1), 26. https://doi.org/10.1007/s10648-024-09871-4

Kohler, B., & Wacker, A. (2013). Das Angebots-Nutzungs-Modell [The offer-use model]. *Die Deutsche Schule*, *105*(3), 241–257.

Komarraju, M., Karau, S. J., & Schmeck, R. R. (2009). Role of the Big Five personality traits in predicting college students' academic motivation and achievement. *Learning and Individual Differences*, *19*(1), 47–52. https://doi.org/10.1016/j.lindif.2008.07.001

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, *49*(4), 241–253. https://doi.org/10.3102/0013189X20912798

Krajcik, J., & Delen, I. (2017). How to support learners in developing usable and lasting knowledge of STEM. *International Journal of Education in Mathematics, Science and Technology*, *5*(1), 21. https://doi.org/10.18404/ijemst.16863

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

Lachner, A., Hoogerheide, V., van Gog, T., & Renkl, A. (2022). Learning-by-teaching without audience presence or interaction: When and why does it work? *Educational Psychology Review*, *34*(2), 575–607. https://doi.org/10.1007/s10648-021-09643-4

Lachner, A., Jacob, L., & Hoogerheide, V. (2021). Learning by writing explanations: Is explaining to a fictitious student more effective than self-explaining? *Learning and Instruction*, *74*, 101438. https://doi.org/10.1016/j.learninstruc.2020.101438

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

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

Lakens, D. (2024). When and how to deviate from a preregistration. *Collabra: Psychology*, *10*(1), 117094. https://doi.org/10.1525/collabra.117094

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

Lazarides, R., & Raufelder, D. (2025). Lernen und Motivation [Learning and motivation]. In M. Syring, T. Bohl, A. Gröschner, & A. Scheunpflug (Eds.), *Studienbuch Bildungswissenschaften 1: Grundbegriffe klären und Forschungszugänge eröffnen* (1st ed., Vol. 1, pp. 64–84). Verlag Julius Klinkhardt. https://doi.org/10.36198/9783838562186-66-86

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

Lipowsky, F. (2020). Unterricht [Teaching]. In E. Wild & J. Möller (Eds.), *Pädagogische Psychologie* (pp. 69–118). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-61403-7_4

Lipowsky, F., & Bleck, V. (2019). Was wissen wir über guten Unterrricht? - Ein Update [What do we know about good teaching? – An update]. In U. Steffens & R. Messner (Eds.), *Unterrichtsqualität: Konzepte und Bilanzen gelingenden Lehrens und Lernens. Grundlagen der Qualität von Schule 3* (pp. 219–249). Waxmann Verlag.

Lipowsky, F., & Rzejak, D. (2021). *Fortbildungen für Lehrpersonen wirksam gestalten [Designing effective professional development for teachers]*. Bertelsmann Stiftung.

Little, T. D. (2024). *Longitudinal structural equation modeling* (2nd ed.). Guilford Publications.

Mammadov, S. (2022). Big Five personality traits and academic performance: A meta-analysis. *Journal of Personality*, *90*(2), 222–255. https://doi.org/10.1111/jopy.12663

Mang, J., Ustjanzew, N., Leßke, I., Schiepe-Tiska, A., & Reiss, K. (2019). *PISA 2015 Skalenhandbuch: Dokumentation der Erhebungsinstrumente [PISA 2015 Scale Handbook: Documentation of Survey Instruments]*. Waxmann.

Mang, J., Ustjanzew, N., Schiepe-Tiska, A., Prenzel, M., Sälzer, C., Müller, K., & Gonzaléz Rodríguez, E. (2018). *PISA 2012 Skalenhandbuch: Dokumentation der Erhebungsinstrumente [PISA 2012 Scale Handbook: Documentation of Survey Instruments]*. Waxmann.

Marsh, H. W., Martin, A. J., Yeung, A. S., & Graven, R. G. (2017). Competence self-perceptions. In A. J. Elliot, C. S. Dweck, & D. S. Yeager (Eds.), *Handbook of competence and motivation: Theory and application* (2nd ed., pp. 85–115). The Guilford Press.

Marsh, H. W., Trautwein, U., Lüdtke, O., & Köller, O. (2008). Social comparison and big-fish-little-pond effects on self-concept and other self-belief constructs: Role of generalized and specific others. *Journal of Educational Psychology*, *100*(3), 510–524. https://doi.org/10.1037/0022-0663.100.3.510

Mason, L., Lowe, R., & Tornatora, M. C. (2013). Self-generated drawings for supporting comprehension of a complex animation. *Contemporary Educational Psychology*, *38*(3), 211–224. https://doi.org/10.1016/j.cedpsych.2013.04.001

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

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511811678

Mayer, R. E., Steinhoff, K., Bower, G., & Mars, R. (1995). A generative theory of textbook design: Using annotated illustrations to foster meaningful learning of science text. *Educational Technology Research and Development*, *43*(1), 31–41. https://doi.org/10.1007/BF02300480

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality*, *60*(2), 175–215. https://doi.org/10.1111/j.1467-6494.1992.tb00970.x

McDermott, K. B., Agarwal, P. K., D'Antonio, L., Roediger, H. L. I., & McDaniel, M. A. (2014). Both multiple-choice and short-answer quizzes enhance later exam performance in middle and high school classes. *Journal of Experimental Psychology: Applied*, *20*(1), 3–21. https://doi.org/10.1037/xap0000004

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, *14*(1), 1–43. https://doi.org/10.1207/s1532690xci1401_1

McNamara, D. S., & Scott, J. L. (1999). Training self explanation and reading strategies. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *43*(21), 1156–1160. https://doi.org/10.1177/154193129904302109

McNeil, N. M., & Uttal, D. H. (2009). Rethinking the use of concrete materials in learning: Perspectives from development and education. *Child Development Perspectives*, *3*(3), 137–139. https://doi.org/10.1111/j.1750-8606.2009.00093.x

Meyer, J., Jansen, T., Hübner, N., & Lüdtke, O. (2023). Disentangling the association between the big five personality traits and student achievement: Meta-analytic evidence on the role of domain specificity and achievement measures. *Educational Psychology Review*, *35*(1), 12. https://doi.org/10.1007/s10648-023-09736-2

Mislevy, R. J. (1995). Test theory and language-learning assessment. *Language Testing*, *12*(3), 341–369. https://doi.org/10.1177/026553229501200305

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

Moreno, R., & Mayer, R. (2007). Interactive multimodal learning environments. *Educational Psychology Review*, *19*(3), 309–326. https://doi.org/10.1007/s10648-007-9047-2

Morphew, J. W. (2021). Changes in metacognitive monitoring accuracy in an introductory physics course. *Metacognition and Learning*, *16*(1), 89–111. https://doi.org/10.1007/s11409-020-09239-3

Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *TIMSS 2019 international results in mathematics and science*. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

Newman, D. A. (2003). Longitudinal modeling with randomly and systematically missing data: A simulation of ad hoc, maximum likelihood, and multiple imputation techniques. *Organizational Research Methods*, *6*(3), 328–362. https://doi.org/10.1177/1094428103254673

Novack, M., & Goldin-Meadow, S. (2015). Learning from gesture: How our hands change our minds. *Educational Psychology Review*, *27*(3), 405–412. https://doi.org/10.1007/s10648-015-9325-3

Obergassel, N., Renkl, A., Endres, T., Nückles, M., Carpenter, S. K., & Roelle, J. (2025). Combining generative tasks and retrieval tasks. *Journal of Educational Psychology*. https://doi.org/10.1037/edu0000949

O'Day, G. M., & Karpicke, J. D. (2021). Comparing and combining retrieval practice and concept mapping. *Journal of Educational Psychology*, *113*(5), 986–997. https://doi.org/10.1037/edu0000486

OECD. (2016). *PISA 2015 results: Volume I: Excellence and equity in education*. OECD Publishing. https://doi.org/10.1787/9789264266490-en

OECD. (2019). *The OECD Learning Compass 2030*. https://www.oecd.org/content/dam/oecd/en/about/projects/edu/education-2040/concept-notes/OECD_Learning_Compass_2030_concept_note.pdf

OECD. (2023). *PISA 2022 results: Volume I: The state of learning and equity in education*. OECD Publishing. https://doi.org/10.1787/53f23881-en

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

Palinscar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, *1*(2), 117–175. https://doi.org/10.1207/s1532690xci0102_1

Pan, S. C., Dunlosky, J., Xu, K. M., & Ouwehand, K. (2024). Emerging and future directions in test-enhanced learning research. *Educational Psychology Review*, *36*(1), 20. https://doi.org/10.1007/s10648-024-09857-2

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*, 81–95. https://doi.org/10.1080/1071441950170102

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

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

Pietsch, M. (2010). Evaluation von Unterrichtsstandards [Evaluation of teaching standards]. *Zeitschrift für Erziehungswissenschaft*, *13*(1), 121–148. https://doi.org/10.1007/s11618-010-0113-z

Pintrich, P. R. (2003). Motivation and classroom learning. In I. B. Weiner (Ed.), *Handbook of psychology* (1st ed., pp. 103–122). Wiley. https://doi.org/10.1002/0471264385.wei0706

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, *82*(1), 33–40. https://doi.org/10.1037/0022-0663.82.1.33

Ploetzner, R., & Fillisch, B. (2017). Not the silver bullet: Learner-generated drawings make it difficult to understand broader spatiotemporal structures in complex animations. *Learning and Instruction*, *47*, 13–24. https://doi.org/10.1016/j.learninstruc.2016.10.002

Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin*, *135*(2), 322–338. https://doi.org/10.1037/a0014996

Praetorius, A.-K., & Gräsel, C. (2021). Noch immer auf der Suche nach dem heiligen Gral: Wie generisch oder fachspezifisch sind Dimensionen der Unterrichtsqualität? [Still in search of the holy grail: How generic or subject-specific are dimensions of teaching quality?]. *Unterrichtswissenschaft*, *49*(2), 167–188. https://doi.org/10.1007/s42010-021-00119-6

Praetorius, A.-K., Rogh, W., & Kleickmann, T. (2020). Blinde Flecken des Modells der drei Basisdimensionen von Unterrichtsqualität? Das Modell im Spiegel einer internationalen Synthese von Merkmalen der Unterrichtsqualität [Blind spots of the model of the three basic dimensions of teaching quality? The model in light of an international synthesis of features of teaching quality]. *Unterrichtswissenschaft*, *48*(3), 303–318. https://doi.org/10.1007/s42010-020-00072-w

Prinz, A., Golke, S., & Wittwer, J. (2018). The double curse of misconceptions: Misconceptions impair not only text comprehension but also metacomprehension in the domain of statistics. *Instructional Science*, *46*(5), 723–765. https://doi.org/10.1007/s11251-018-9452-6

Prinz, A., Golke, S., & Wittwer, J. (2020a). How accurately can learners discriminate their comprehension of texts? A comprehensive meta-analysis on relative metacomprehension accuracy and influencing factors. *Educational Research Review*, *31*, 100358. https://doi.org/10.1016/j.edurev.2020.100358

Prinz, A., Golke, S., & Wittwer, J. (2020b). To what extent do situation-model-approach interventions improve relative metacomprehension accuracy? Meta-analytic insights. *Educational Psychology Review*, *32*(4), 917–949. https://doi.org/10.1007/s10648-020-09558-6

R Core Team. (2024). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

R Core Team. (2025). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Rasco, R. W., Tennyson, R. D., & Boutwell, R. C. (1975). Imagery instructions and drawings in learning prose. *Journal of Educational Psychology*, *67*(2), 188–192. https://doi.org/10.1037/h0077014

Rau, M. A. (2017). Conditions for the effectiveness of multiple visual representations in enhancing STEM learning. *Educational Psychology Review*, *29*(4), 717–761. https://doi.org/10.1007/s10648-016-9365-3

Raudenbush, S. W. (2008). Advancing educational policy by advancing research on instruction. *American Educational Research Journal*, *45*(1), 206–230. https://doi.org/10.3102/0002831207312905

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). SAGE.

Rawson, K. A., & Dunlosky, J. (2022). Successive relearning: An underexplored but potent technique for obtaining and maintaining knowledge. *Current Directions in Psychological Science*, *31*(4), 362–368. https://doi.org/10.1177/09637214221100484

Renkl, A. (2009). Lehren und Lernen. In R. Tippelt & B. Schmidt-Hertha (Eds.), *Handbuch Bildungsforschung* (2., überarbeitete und erw. Aufl, pp. 737–751). VS, Verlag für Sozialwissenschaften.

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

Ribosa, J., & Duran, D. (2022). Do students learn what they teach when generating teaching materials for others? A meta-analysis through the lens of learning by teaching. *Educational Research Review*, *37*, 100475. https://doi.org/10.1016/j.edurev.2022.100475

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin*, *138*(2), 353–387. https://doi.org/10.1037/a0026838

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*, *38*(4), 1141–1159. https://doi.org/10.1111/jcal.12673

Richter, J., Scheiter, K., & Eitel, A. (2018). Signaling text–picture relations in multimedia learning: The influence of prior knowledge. *Journal of Educational Psychology*, *110*(4), 544–560. https://doi.org/10.1037/edu0000220

Richter, T., Berger, R., Ebersbach, M., Eitel, A., Endres, T., Borromeo Ferri, R., Hänze, M., Lachner, A., Leutner, D., Lipowsky, F., Nemeth, L., Renkl, A., Roelle, J., Rummer, R., Scheiter, K., Schweppe, J., von Aufschnaiter, C., & Vorholzer, A. (2022). How to promote lasting learning in schools. *Zeitschrift Für Entwicklungspsychologie Und Pädagogische Psychologie*, *54*(4), 135–141. https://doi.org/10.1026/0049-8637/a000258

Richter, T., & Maier, J. (2017). Comprehension of multiple documents with conflicting information: A two-step model of validation. *Educational Psychologist*, *52*(3), 148–166. https://doi.org/10.1080/00461520.2017.1322968

Rieger, S., Göllner, R., Spengler, M., Trautwein, U., Nagengast, B., & Roberts, B. W. (2022). The persistence of students' academic effort: The unique and combined effects of conscientiousness and individual interest. *Learning and Instruction*, *80*, 101613. https://doi.org/10.1016/j.learninstruc.2022.101613

Roberts, B. W., Lejuez, C., Krueger, R. F., Richards, J. M., & Hill, P. L. (2014). What is conscientiousness and how can it be assessed? *Developmental Psychology*, *50*(5), 1315–1330. https://doi.org/10.1037/a0031109

Roediger, H. L., Agarwal, P. K., McDaniel, M. A., & McDermott, K. B. (2011). Test-enhanced learning in the classroom: Long-term improvements from quizzing. *Journal of Experimental Psychology: Applied*, *17*(4), 382–395. https://doi.org/10.1037/a0026252

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15*(1), 20–27. https://doi.org/10.1016/j.tics.2010.09.003

Roediger, H. L., Jacoby, J. D., & McDermott, K. B. (1996). Misinformation effects in recall: Creating false memories through repeated retrieval. *Journal of Memory and Language*, *35*(2), 300–318. https://doi.org/10.1006/jmla.1996.0017

Roelle, J., Endres, T., Abel, R., Obergassel, N., Nückles, M., & Renkl, A. (2023). Happy together? On the relationship between research on retrieval practice and generative learning using the case of follow-up learning tasks. *Educational Psychology Review*, *35*(4), 102. https://doi.org/10.1007/s10648-023-09810-9

Roelle, J., Froese, L., Krebs, R., Obergassel, N., & Waldeyer, J. (2022). Sequence matters! Retrieval practice before generative learning is more effective than the reverse order.

Learning and Instruction, *80*, 101634. https://doi.org/10.1016/j.learninstruc.2022.101634

Roelle, J., Lachner, A., & Heitmann, S. (2023). *Lernen: Theorien und Techniken [Learning: Theories and techniques]*. UTB.

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

Roelle, J., Schweppe, J., Endres, T., Lachner, A., Aufschnaiter, C. V., Renkl, A., Eitel, A., Leutner, D., Rummer, R., Scheiter, K., & Vorholzer, A. (2022). Combining retrieval practice and generative learning in educational contexts: Promises and challenges. *Zeitschrift Für Entwicklungspsychologie Und Pädagogische Psychologie*, *54*(4), 142–150. https://doi.org/10.1026/0049-8637/a000261

Roelle, J., Schweppe, J., Endres, T., Lachner, A., von Aufschnaiter, C., Renkl, A., Eitel, A., Leutner, D., Rummer, R., Scheiter, K., & Vorholzer, A. (2022). Combining retrieval practice and generative learning in educational contexts: Promises and challenges. *Zeitschrift Für Entwicklungspsychologie Und Pädagogische Psychologie*, *54*(4), 142–150. https://doi.org/10.1026/0049-8637/a000261

Rolff, H.-G. (2018). Bildungsreform als Schulentwicklung [Educational reform as school development]. In H. Barz (Ed.), *Handbuch Bildungsreform und Reformpädagogik* (pp. 595–603). Springer VS. https://doi.org/10.1007/978-3-658-07491-3

Roscoe, R. D. (2014). Self-monitoring and knowledge-building in learning by teaching. *Instructional Science*, *42*(3), 327–351. https://doi.org/10.1007/s11251-013-9283-4

Rosenthal, R., & Rosnow, R. L. (1985). *Contrast analysis: Focused comparisons in the analysis of variance*. Cambridge University Press.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, *140*(6), 1432–1463. https://doi.org/10.1037/a0037559

Rzejak, D., Lipowsky, F., & Bleck, V. (2020). Synopse bedeutsamer Merkmale von Lehrkräftefortbildungen. Wirkungsebenen und konzeptionelle Aspekte. *Journal für LehrerInnenbildung*, *4*. https://doi.org/10.35468/jlb-04-2020-01

Sarama, J., & Clements, D. H. (2009). “Concrete” computer manipulatives in mathematics education. *Child Development Perspectives*, *3*(3), 145–150. https://doi.org/10.1111/j.1750-8606.2009.00095.x

Scheiter, K., Schleinschok, K., & Ainsworth, S. (2017). Why sketching may aid learning from science texts: Contrasting sketching with written explanations. *Topics in Cognitive Science*, *9*(4), 866–882. https://doi.org/10.1111/tops.12261

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

Schmeck, A., Mayer, R. E., Opfermann, M., Pfeiffer, V., & Leutner, D. (2014). Drawing pictures during learning from scientific text: Testing the generative drawing effect and the prognostic drawing effect. *Contemporary Educational Psychology*, *39*(4), 275–286. https://doi.org/10.1016/j.cedpsych.2014.07.003

Schmidgall, S. P., Eitel, A., & Scheiter, K. (2019). Why do learners who draw perform well? Investigating the role of visualization, generation and externalization in learner-generated drawing. *Learning and Instruction*, *60*, 138–153. https://doi.org/10.1016/j.learninstruc.2018.01.006

Schneider, M., Rittle-Johnson, B., & Star, J. R. (2011). Relations among conceptual knowledge, procedural knowledge, and procedural flexibility in two samples differing in prior knowledge. *Developmental Psychology*, *47*(6), 1525–1538. https://doi.org/10.1037/a0024997

Schraw, G. (1994). The effect of metacognitive knowledge on local and global monitoring. *Contemporary Educational Psychology*, *19*(2), 143–154. https://doi.org/10.1006/ceps.1994.1013

Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and Learning*, *4*(1), 33–45. https://doi.org/10.1007/s11409-008-9031-3

Schwamborn, A., Mayer, R. E., Thillmann, H., Leopold, C., & Leutner, D. (2010). Drawing as a generative activity and drawing as a prognostic activity. *Journal of Educational Psychology*, *102*(4), 872–879. https://doi.org/10.1037/a0019640

Schwamborn, A., Thillmann, H., Opfermann, M., & Leutner, D. (2011). Cognitive load and instructionally supported learning with provided and learner-generated visualizations. *Computers in Human Behavior*, *27*(1), 89–93. https://doi.org/10.1016/j.chb.2010.05.028

Seiz, J., Decristan, J., Kunter, M., & Baumert, J. (2016). Differenzielle Effekte von Klassenführung und Unterstützung für Schülerinnen und Schüler mit Migrationshintergrund [Differential effects of classroom management and support for students with a migration background]. *Zeitschrift für Pädagogische Psychologie*, *30*(4), 237–249. https://doi.org/10.1024/1010-0652/a000186

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research*, *46*(3), 407–441. https://doi.org/10.3102/00346543046003407

Sibley, L., Fiorella, L., & Lachner, A. (2022). It's better when I see it: Students benefit more from open-book than closed-book teaching. *Applied Cognitive Psychology*, *36*(6), 1347–1355. https://doi.org/10.1002/acp.4017

Sibley, L., Russ, H., Ahmad, G., Baumgärtner, B., Bräutigam, D., Brümmer, S., Bussmann, H., Erb, N., Evans, L., Fischer, S., Gradl, L., Guddemi, R., Hauptmann, K.-B., Hieke, J., Hilsdorf, S., Högerle, F., Hoppe-Brixner, B., Jeong, W., Karl, S., … Lachner, A. (2024). Does technology-based non-interactive teaching enhance students' learning in the classroom? *Computers and Education Open*, *7*, 100233. https://doi.org/10.1016/j.caeo.2024.100233

Snow, R. E. (1991). Aptitude-treatment interaction as a framework for research on individual differences in psychotherapy. *Journal of Consulting and Clinical Psychology*, *59*(2), 205–216. https://doi.org/10.1037/0022-006X.59.2.205

Song, J., Gaspard, H., Nagengast, B., & Trautwein, U. (2020). The Conscientiousness × Interest Compensation (CONIC) model: Generalizability across domains, outcomes, and predictors. *Journal of Educational Psychology*, *112*(2), 271–287. https://doi.org/10.1037/edu0000379

Spielmann, J., Yoon, H. J. R., Ayoub, M., Chen, Y., Eckland, N. S., Trautwein, U., Zheng, A., & Roberts, B. W. (2022). An in-depth review of conscientiousness and educational issues. *Educational Psychology Review*, *34*(4), 2745–2781. https://doi.org/10.1007/s10648-022-09693-2

Sweller, J. (2024). Cognitive load theory and individual differences. *Learning and Individual Differences*, *110*, 102423. https://doi.org/10.1016/j.lindif.2024.102423

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer. https://doi.org/10.1007/978-1-4419-8126-4

Syring, M., Bohl, T., & Lachner, A. (2022). Digitalisierung in der Schule: Vorschlag eines systematisierenden Rahmenmodells aus schulpädagogischer Perspektive

[Digitalization in schools: A proposed framework model from an educational pedagogy perspective]. *Zeitschrift für Bildungsforschung*, *12*(3), 615–630. https://doi.org/10.1007/s35834-022-00340-y

Täschner, J., Holzberger, D., & Hillmayr, D. (2021). Elternbeteiligung als Potenzial zur Förderung des schulischen Erfolgs. Ein Second-Order-Review [Parental involvement as a potential for promoting academic success. A second order review]. *DDS – Die Deutsche Schule*, *2021*(3), 302–317. https://doi.org/10.31244/dds.2021.03.05

Terhart, E. (2019). *Didaktik: Eine Einführung [Didactics: An introduction]* (1. Auflage). Reclam Verlag.

Urhahne, D., & Wijnia, L. (2023). Theories of motivation in education: An integrative framework. *Educational Psychology Review*, *35*(2), 45. https://doi.org/10.1007/s10648-023-09767-9

Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational Psychologist*, *39*(2), 111–133. https://doi.org/10.1207/s15326985ep3902_3

Van Essen, G., & Hamaker, C. (1990). Using self-generated drawings to solve arithmetic word problems. *The Journal of Educational Research*, *83*(6), 301–312. https://doi.org/10.1080/00220671.1990.10885976

Van Gog, T., Kester, L., Dirkx, K., Hoogerheide, V., Boerboom, J., & Verkoeijen, P. P. J. L. (2015). Testing after worked example study does not enhance delayed problem-solving performance compared to restudy. *Educational Psychology Review*, *27*(2), 265–289. https://doi.org/10.1007/s10648-015-9297-3

Van Meter, P. (2001). Drawing construction as a strategy for learning from text. *Journal of Educational Psychology*, *93*(1), 129–140. https://doi.org/10.1037/0022-0663.93.1.129

Van Meter, P., Aleksic, M., Schwartz, A., & Garner, J. (2006). Learner-generated drawing as a strategy for learning from content area text. *Contemporary Educational Psychology*, *31*(2), 142–166. https://doi.org/10.1016/j.cedpsych.2005.04.001

Van Meter, P., & Firetto, C. M. (2013). Cognitive model of drawing construction: Learning through the construction of drawings. In G. Schraw, M. T. McCrudden, & D. Robinson (Eds.), *Learning through visual displays* (pp. 247–280). IAP.

Van Meter, P., & Garner, J. (2005). The promise and practice of learner-generated drawing: Literature review and synthesis. *Educational Psychology Review*, *17*(4), 285–325. https://doi.org/10.1007/s10648-005-8136-3

VanLehn, K. (1996). Cognitive skill acquisition. *Annual Review of Psychology*, *47*(1), 513–539. https://doi.org/10.1146/annurev.psych.47.1.513

Vieluf, S., Praetorius, A.-K., Rakoczy, K., Kleinknecht, M., & Pietsch, M. (2020). Angebots-Nutzungs-Modelle der Wirkweise des Unterrichts. Ein kritischer Vergleich verschiedener Modellvarianten [Offer-use models of how teaching works: A critical comparison of different model variants]. *Zeitschrift für Pädagogik*, *66*(Beiheft), 63–80. https://doi.org/10.25656/01:25864

Wagner, S., Sibley, L., Weiler, D., Burde, J.-P., Scheiter, K., & Lachner, A. (2024). The more, the better? Learning with feedback and instruction. *Learning and Instruction*, *89*, 101844. https://doi.org/10.1016/j.learninstruc.2023.101844

Waldeyer, J., Dicke, T., Fleischer, J., Guo, J., Trentepohl, S., Wirth, J., & Leutner, D. (2022). A moderated mediation analysis of conscientiousness, time management strategies, effort regulation strategies, and university students' performance. *Learning and Individual Differences*, *100*, 102228. https://doi.org/10.1016/j.lindif.2022.102228

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

Wang, H., Liu, Y., Wang, Z., & Wang, T. (2023). The influences of the Big Five personality traits on academic achievements: Chain mediating effect based on major identity and self-efficacy. *Frontiers in Psychology*, *14*, 1065554. https://doi.org/10.3389/fpsyg.2023.1065554

Wang, J., & Wang, X. (2020). *Structural equation modeling: Applications using mplus* (2nd ed.). John Wiley & Sons.

Wang, Y., & Zhang, W. (2024). The relationship between college students' learning engagement and academic self-efficacy: A moderated mediation model. *Frontiers in Psychology*, *15*, 1425172. https://doi.org/10.3389/fpsyg.2024.1425172

Wiens, S., & Nilsson, M. E. (2017). Performing contrast analysis in factorial designs: From NHST to confidence intervals and beyond. *Educational and Psychological Measurement*, *77*(4), 690–715. https://doi.org/10.1177/0013164416668950

Wittrock, M. C. (1989). Generative processes of comprehension. *Educational Psychologist*, *24*(4), 345–376. https://doi.org/10.1207/s15326985ep2404_2

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

Wolf, L. F., & Smith, J. K. (1995). The consequence of consequence: Motivation, anxiety, and test performance. *Applied Measurement in Education*, *8*(3), 227. https://doi.org/10.1207/s15324818ame0803_3

Wörner, S., Becker, S., Küchemann, S., Scheiter, K., & Kuhn, J. (2022). Development and validation of the ray optics in converging lenses concept inventory. *Physical Review Physics Education Research*, *18*(2), 020131. https://doi.org/10.1103/PhysRevPhysEducRes.18.020131

Wu, S. P. W., & Rau, M. A. (2019). How students learn content in Science, Technology, Engineering, and Mathematics (STEM) through drawing activities. *Educational Psychology Review*, *31*(1), 87–120. https://doi.org/10.1007/s10648-019-09467-3

Yang, C., Luo, L., Vadillo, M. A., Yu, R., & Shanks, D. R. (2021). Testing (quizzing) boosts classroom learning: A systematic and meta-analytic review. *Psychological Bulletin*, *147*(4), 399–435. https://doi.org/10.1037/bul0000309

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

Zhuang, L., Wang, J., Xiong, B., Bian, C., Hao, L., Bayley, P. J., & Qin, S. (2022). Rapid neural reorganization during retrieval practice predicts subsequent long-term retention and false memory. *Nature Human Behaviour*, *6*(1), 134–145. https://doi.org/10.1038/s41562-021-01188-4

8

APPENDIX

8 Appendix

8.1 Appendix Study 1

Appendix A

Overview of the Overall Dataset

Variable	Number/type of items	Used in present paper
Demographical data		
Age	1 item, rated from 10 to 17 years	×
Sex	1 item, closed answer format	×
Native language	1 item, open answer format	×
School type & level	1 item, closed answer format	–
Grade in physics, last report	1 item, rated from 1 to 6	–
Graduation of parents	1 item, closed answer format	–
Students' prerequisites		
Interest in physics	4 items, rated on a 4-point Likert scale	×
Physics work ethic	4 items, rated on a 4-point Likert scale	×
Academic self-concept in physics	4 items, rated on a 4-point Likert scale	×
ICT interest	4 items, rated on a 4-point Likert scale	×
Experience with touch devices	4 items, rated on a 5-point Likert scale	–
Monitoring rating	1 item, rated from 0 to 30 points	×
Prior knowledge		
Cognitive	15 items, multiple-choice-format	×
Metacognitive[a]		×
Big five personality trait domains		
Extraversion	3 items, rated on a 5-point Likert scale	–
Compatibility	3 items, rated on a 5-point Likert scale	–
Conscientiousness	3 items, rated on a 5-point Likert scale	–
Neuroticism	3 items, rated on a 5-point Likert scale	–
Openness	3 items, rated on a 5-point Likert scale	–
Perceived ratings regarding the teaching unit		
Cognitive load		
Active cognitive load	1 item, rated on a 9-point Likert scale	×
Passive cognitive load	1 item, rated on a 9-point Likert scale	×
Affect		
Arousal	1 item, rated on a 9-point Likert scale	×
Mood	1 item, rated on a 9-point Likert scale	×
Teaching quality		
Cognitive activation	6 items, rated on a 4-point Likert scale	×
Disturbances	3 items, rated on a 4-point Likert scale	×
Teacher monitoring	5 items, rated on a 4-point Likert scale	×
Teacher support	4 items, rated on a 4-point Likert scale	×
Characteristics of students' explanations		
Completeness	1 coded item, range 0 to 10 points	×

Variable	Number/type of items	Used in present paper
Elaboration	1 coded item, 1 point per elaboration	×
Correctness	1 coded item, percentage value	×
Characteristics of students' drawings		
Completeness	1 coded item, range 0 to 10 points	×
Elaboration	1 coded item, 1 point per elaboration	×
Correctness	1 coded item, percentage value	×
Perceived ratings regarding the learning activities		
Cognitive load		
Active cognitive load	1 item, rated on a 9-point Likert scale	–
Passive cognitive load	1 item, rated on a 9-point Likert scale	–
Task-specific motivation		
Task interest	2 items, rated on a 4-point Likert scale	×
Task enjoyment	2 items, rated on a 4-point Likert scale	×
Affect		
Arousal	1 item, rated on a 9-point Likert scale	–
Mood	1 item, rated on a 9-point Likert scale	–
Perceived rating immediate after learning activity		
Monitoring rating	1 item, rated from 0 to 30 points	×
Immediate learning outcomes		
Cognitive	15 items, multiple-choice-format	×
Metacognitive[a]		×
Perceived rating delayed after learning activity		
Monitoring rating	1 item, rated from 0 to 30 points	×
Lasting learning outcomes		
Cognitive	15 items, multiple choice format	×
Metacognitive[a]		×

Note. All variables are listed in the correct chronological order. [a]Metacognitive learning outcomes are each quantified as monitoring accuracy, which is calculated based on the absolute differences between students' estimated and actual performance without negative values.

Appendix B

Correlations Separated by Experimental Condition

Variable	1	2	3	4	5
1. Prior cognitive knowledge	–				
2. Prior metacognitive knowledge	(-.52*|-.50*| -.51*|-.46*)	–			
3. Students' explanations completeness[a]	(-|.13| .17|-.02)	(-|.04| -.07|.11)	–		
4. Students' explanations elaboration[a]	(-|.31*| .22*|.17)	(-|-.07| -.08|-.08)	(-|.36*| .39*|.22*)	–	
5. Students' explanations correctness[a]	(-|-.03| .10|.08)	(-|.08| -.09|-.09)	(-|.39*| .40*|.10)	(-|.20*| .12|.06)	–
6. Students' drawings completeness[a]	(-|-| -|.04)	(-|-| -|.00)	(-|-| -|.23*)	(-|-| -|.01)	(-|-| -|.11)
7. Students' drawings elaboration[a]	(-|-| -|.14)	(-|-| -|-.05)	(-|-| -|-.03)	(-|-| -|.13)	(-|-| -|.20*)
8. Students' drawings correctness[a]	(-|-| -|.07)	(-|-| -|.02)	(-|-| -|.20*)	(-|-| -|-.02)	(-|-| -|.27*)
9. Task interest[a]	(-.07|-.04| .02|.07)	(.15|.09| -.05|.16)	(-|.36*| .35*|.27*)	(-|.16| .21*|.16)	(-|.16| .14|.02)
10. Task enjoyment[a]	(-.11|-.08| .01|.11)	(.22*|.19*| -.01|.10)	(-|.26*| .32*|.14)	(-|.11| .22*|.07)	(-|.09| .13|.04)
11. Immediate cognitive learning outcome	(.32*|.35*| .29*|.35*)	(-.16|-.05| -.20*|-.09)	(-|.37*| .30*|.36*)	(-|.37*| .27*|.14)	(-|.16| .32*|.20*)
12. Immediate metacognitive learning outcome	(-.07|-.03| -.05|-.06)	(.13|-.04| .11|-.09)	(-|.04| -.20*|.10)	(-|.04| .04|.08)	(-|.02| -.13|.01)
13. Lasting cognitive learning outcome	(.44*|.43*| .39*|.54*)	(-.23*|-.16| -.27*|-.20*)	(-|.25*| .29*|.22*)	(-|.38*| .17|.08)	(-|-.01| .14|.15)
14. Lasting metacognitive learning outcome	(-.06|-.12| .14|-.15)	(.14|.16| .01|.11)	(-|-.21*| -.03|.14)	(-|-.07| .03|-.09)	(-|-.12| .03|.03)

Variable	6	7	8	9	10
1. Prior cognitive knowledge					
2. Prior metacognitive knowledge					
3. Students' explanations completeness[a]					
4. Students' explanations elaboration[a]					
5. Students' explanations correctness[a]					
6. Students' drawings completeness[a]	–				
7. Students' drawings elaboration[a]	(-|-|-|.14)	–			
8. Students' drawings correctness[a]	(-|-|-|.40*)	(-|-|-|.09)	–		
9. Task interest[a]	(-|-|-|.09)	(-|-|-|.18*)	(-|-|-|.01)	–	
10. Task enjoyment[a]	(-|-|-|.10)	(-|-|-|.19*)	(-|-|-|.03)	(.82*|.81*|.81*|.68*)	–
11. Immediate cognitive learning outcome	(-|-|-|.16*)	(-|-|-|.12)	(-|-|-|.18*)	(.06|.29*|.17*|.17*)	(.08|.24*|.19*|.17*)
12. Immediate metacognitive learning outcome	(-|-|-|-.02)	(-|-|-|.01)	(-|-|-|.05)	(-.02|-.05|-.11|.05)	(-.09|.00|-.11|.05)
13. Lasting cognitive learning outcome	(-|-|-|.08)	(-|-|-|.13)	(-|-|-|.15)	(.10|.19*|.04|.10)	(.09|.14|.07|.10)
14. Lasting metacognitive learning outcome	(-|-|-|.10)	(-|-|-|.03)	(-|-|-|.14)	(-.05|-.25*|-.05|-.06)	(-.10|-.16|-.03|-.02)

Variable	11	12	13
1. Prior cognitive knowledge			
2. Prior metacognitive knowledge			
3. Students' explanations completeness[a]			
4. Students' explanations elaboration[a]			
5. Students' explanations correctness[a]			
6. Students' drawings completeness[a]			
7. Students' drawings elaboration[a]			
8. Students' drawings correctness[a]			
9. Task interest[a]			
10. Task enjoyment[a]			
11. Immediate cognitive learning outcome	–		
12. Immediate metacognitive learning outcome	(-.08|-.12| .14|.26*)	–	
13. Lasting cognitive learning outcome	(.58*|.65*| .61*|.65*)	(.01|-.14| .12|.06)	–
14. Lasting metacognitive learning outcome	(-.10|-.21*| -.12|.01)	(.27*|.22*| .26*|.17*)	(-.17|-.22*| -.11|-.04)

Note. The data in this table are based on the raw data. Numbers in brackets represent the correlations separated for experimental conditions: first = restudy; second = teaching-only; third = teaching + visualization; fourth = teaching + drawing.
[a] Variables regarding the learning activity (restudy, teaching-only, teaching + visualization, teaching + drawing).
[*] $p < .050$.

8.2 Appendix Study 2

Appendix A

Overview of the Schools, Classes, and Final Number of Participants in the Study

School	Class	Class size	Number of participants[a]	Number of excluded participants[b]	Final number of participants[c]
1	1	27	27	1	26
	2	28	28	3	25
	3	24	23	0	23
	4	26	24	4	20
2	5	27	27	3	24
	6	27	27	3	24
3	7	25	25	0	25
	8	25	25	3	22
	9	24	22	3	19
	10	25	25	2	23
4	11	22	22	1	21
	12	22	22	3	19
	13	24	23	0	23
	14	25	25	2	23

Note. [a]Students who provided written consent by their legal guardians to voluntarily participate in the study (N = 345). [b]We excluded data from participants who did not attend the main part of the study, as it included the topic-related teaching unit, the interventions, and the immediate posttest (n = 28). [c]This resulted in a final sample size of N = 317.

Appendix B

Overview of the Overall Dataset

Variable	Number/type of items	Used in present paper
Demographical data		
Age	1 item, rated from 10 to 17 years	×
Sex	1 item, closed answer format	×
Native language	1 item, open answer format	×
Schooltype & level	1 item, closed answer format	–
Grade in physics, last report	1 item, rated from 1 to 6	–
Graduation of parents	1 item, closed answer format	–
Students' prerequisites		
Interest in physics	4 items, rated on a 4-point Likert scale	–
Physics work ethic	4 items, rated on a 4-point Likert scale	×
Academic self-concept in physics	4 items, rated on a 4-point Likert scale	×
ICT interest	4 items, rated on a 4-point Likert scale	–
Experience with touch devices	4 items, rated on a 5-point Likert scale	–
Monitoring rating	1 item, rated from 0 to 30 points	×
Prior		
Conceptual knowledge	15 items, multiple-choice-format	×
Monitoring accuracy[a]		×
Big five personality trait domains		
Extraversion	3 items, rated on a 5-point Likert scale	–
Compatibility	3 items, rated on a 5-point Likert scale	–
Conscientiousness	3 items, rated on a 5-point Likert scale	–
Neuroticism	3 items, rated on a 5-point Likert scale	–
Openness	3 items, rated on a 5-point Likert scale	–
Perceived ratings after learning phase		
Active cognitive load	1 item, rated on a 9-point Likert scale	–
Passive cognitive load	1 item, rated on a 9-point Likert scale	–
Arousal	1 item, rated on a 9-point Likert scale	–
Mood	1 item, rated on a 9-point Likert scale	–
Teaching quality		
Cognitive activation	6 items, rated on a 4-point Likert scale	–
Disturbances	3 items, rated on a 4-point Likert scale	–
Teacher monitoring	5 items, rated on a 4-point Likert scale	–
Teacher support	4 items, rated on a 4-point Likert scale	–
Completion of students' experimentation worksheets	1 coded item, 0 = not fully and correctly completed, 1 = fully and correctly completed	×
Characteristics of students' explanations		
Completeness	1 coded item, range 0 to 10 points	×
Elaboration	1 coded item, 1 point per elaboration	×
Correctness	1 coded item, percentage value	×
Characteristics of students' drawings		
Completeness	1 coded item, range 0 to 10 points	×
Elaboration	1 coded item, 1 point per elaboration	×
Correctness	1 coded item, percentage value	×
Characteristics of students' restudy notes		
Completeness	1 coded item, range 0 to 10 points	×
Elaboration	1 coded item, 1 point per elaboration	×
Correctness	1 coded item, percentage value	×
Perceived ratings immediately after learning activity		
Active cognitive load	1 item, rated on a 9-point Likert scale	–
Passive cognitive load	1 item, rated on a 9-point Likert scale	–
Task interest	2 items, rated on a 4-point Likert scale	–
Task enjoyment	2 items, rated on a 4-point Likert scale	–

Variable	Number/type of items	Used in present paper
Arousal	1 item, rated on a 9-point Likert scale	–
Mood	1 item, rated on a 9-point Likert scale	–
Monitoring rating	1 item, rated from 0 to 30 points	×
Immediate		
Conceptual knowledge	15 items, multiple-choice-format	×
Monitoring accuracy[a]		×
Perceived rating delayed after learning activity		
Monitoring rating	1 item, rated from 0 to 30 points	×
Lasting		
Conceptual knowledge	15 items, multiple choice format	×
Monitoring accuracy[a]		×

Note. All variables are listed in the correct chronological order.

[a]Monitoring accuracy is calculated based on the differences between students' estimated and actual performance.

Appendix C

Summary of the Contrast Analyses

Variable	β	*SE*	*t*	*p*
Conceptual knowledge				
Immediate				
Generation contrast[a]	**0.11**	**0.04**	**2.83**	**.005**
Drawing contrast[b]	0.10	0.07	1.44	.150
Distribution contrast[c]	-0.06	0.05	-1.10	.272
Prior cognitive knowledge	**0.23**	**0.06**	**3.73**	**<.001**
Interaction (Generation × Distribution)	**-0.09**	**0.04**	**-2.41**	**.017**
Interaction (Drawing × Distribution)	-0.02	0.07	-0.34	.737
Lasting				
Generation contrast[a]	0.06	0.04	1.50	.134
Drawing contrast[b]	0.10	0.07	1.34	.181
Distribution contrast[c]	-0.04	0.06	-0.68	.495
Prior cognitive knowledge	**0.24**	**0.06**	**3.99**	**<.001**
Interaction (Generation × Distribution)	-0.04	0.04	-0.86	.389
Interaction (Drawing × Distribution)	0.00	0.08	-0.03	.972
Monitoring accuracy				
Immediate				
Generation contrast[a]	**-0.09**	**0.04**	**-2.49**	**.013**
Drawing contrast[b]	-0.06	0.06	-0.94	.349
Distribution contrast[c]	0.02	0.05	0.32	.747
Prior metacognitive knowledge	**0.38**	**0.06**	**6.57**	**<.001**
Interaction (Generation × Distribution)	0.06	0.04	1.69	.093
Interaction (Drawing × Distribution)	0.00	0.06	0.08	.939
Lasting				
Generation contrast[a]	-0.06	0.04	-1.47	.144
Drawing contrast[b]	-0.02	0.08	-0.33	.742
Distribution contrast[c]	0.06	0.06	1.03	.304
Prior metacognitive knowledge	**0.19**	**0.07**	**2.91**	**.004**
Interaction (Generation × Distribution)	0.03	0.04	0.68	.500
Interaction (Drawing × Distribution)	-0.02	0.07	-0.32	.751

Note. Significant results are highlighted in bold letters; $p < .050$.

[a] -2 = restudy, 1 = teaching-only, 1 = teaching + drawing. [b] 0 = restudy, -1 = teaching-only, 1 = teaching + drawing. [c] -1 = after-study, 1 = distributed.

Appendix D

Summary of the Mediation Analyses With Characteristics of Students' Explanations and Restudy Notes Regarding Immediate Conceptual Knowledge

Variable	β	*SE*	*t*	*p*
Completeness				
Effect of generation contrast[a] on completeness (a)	**0.60**	**0.12**	**5.19**	**<.001**
Effect of completeness on immediate conceptual knowledge (b)	**0.25**	**0.06**	**4.44**	**<.001**
Direct effect (c')	0.19	0.12	1.57	.118
Indirect effect (a*b)	**0.15**	**0.04**	**3.37**	**.001**
Total effect (c)	**0.34**	**0.12**	**2.84**	**.005**
Elaboration				
Effect of generation contrast[a] on elaboration (a)	**0.34**	**0.12**	**2.82**	**.005**
Effect of elaboration on immediate conceptual knowledge (b)	0.09	0.06	1.60	.110
Direct effect (c')	**0.31**	**0.12**	**2.56**	**.010**
Indirect effect (a*b)	0.03	0.02	1.39	.164
Total effect (c)	**0.34**	**0.12**	**2.83**	**.005**
Correctness				
Effect of generation contrast[a] on correctness (a)	-0.16	0.12	-1.37	.171
Effect of correctness on immediate conceptual knowledge (b)	**0.13**	**0.06**	**2.41**	**.016**
Direct effect (c')	**0.36**	**0.12**	**3.04**	**.002**
Indirect effect (a*b)	-0.02	0.02	-1.19	.234
Total effect (c)	**0.34**	**0.12**	**2.84**	**.005**

Note. Significant results are highlighted in bold letters; $p < .050$.

[a] -2 = restudy, 1 = teaching-only, 1 = teaching + drawing.

Appendix E

Summary of the Moderation Analyses With Prior Conceptual Knowledge, Academic Self-Concept, and Work Ethic Regarding Immediate and Lasting Conceptual Knowledge

Variable	β	*SE*	*t*	*p*
Immediate conceptual knowledge				
Prior conceptual knowledge as moderator				
Generation contrast[a]	**0.10**	**0.04**	**2.72**	**.007**
Drawing contrast[b]	0.10	0.07	1.43	.154
Distribution contrast[c]	-0.06	0.05	-1.10	.272
Prior conceptual knowledge	**0.23**	**0.06**	**3.62**	**<.001**
Interaction (Generation × Distribution)	**-0.09**	**0.04**	**-2.29**	**.022**
Interaction (Drawing × Distribution)	-0.02	0.07	-0.27	.791
Interaction (Generation × Prior conceptual knowledge)	-0.04	0.05	-0.84	.404
Interaction (Drawing × Prior conceptual knowledge)	0.02	0.07	0.33	.742
Interaction (Distribution × Prior conceptual knowledge)	0.07	0.06	1.04	.299
Interaction (Generation × Distribution × Prior conceptual knowledge)	0.04	0.05	0.95	.345
Interaction (Drawing × Distribution × Prior conceptual knowledge)	0.04	0.07	0.56	.577
Academic self-concept as moderator				
Generation contrast[a]	**0.10**	**0.04**	**2.59**	**.010**
Drawing contrast[b]	0.09	0.07	1.29	.199
Distribution contrast[c]	-0.04	0.05	-0.68	.496
Academic self-concept	**0.22**	**0.06**	**3.88**	**<.001**
Prior conceptual knowledge	**0.22**	**0.06**	**3.75**	**<.001**
Interaction (Generation × Distribution)	**-0.09**	**0.04**	**-2.30**	**.022**
Interaction (Drawing × Distribution)	-0.03	0.07	-0.42	.674
Interaction (Generation × Academic self-concept)	0.02	0.04	0.49	.621
Interaction (Drawing × Academic self-concept)	0.03	0.07	0.35	.726
Interaction (Distribution × Academic self-concept)	-0.01	0.06	-0.18	.859
Interaction (Generation × Distribution × Academic self-concept)	**-0.08**	**0.04**	**-2.23**	**.026**
Interaction (Drawing × Distribution × Academic self-concept)	0.02	0.07	0.28	.780
Work ethic as moderator				
Generation contrast[a]	**0.11**	**0.04**	**2.91**	**.004**
Drawing contrast[b]	0.09	0.07	1.30	.193
Distribution contrast[c]	-0.05	0.05	-0.99	.321
Work ethic	**0.11**	**0.06**	**1.98**	**.048**
Prior conceptual knowledge	**0.23**	**0.06**	**3.84**	**<.001**
Interaction (Generation × Distribution)	**-0.08**	**0.04**	**-2.23**	**.027**
Interaction (Drawing × Distribution)	-0.01	0.07	-0.10	.921
Interaction (Generation × Work ethic)	0.03	0.04	0.90	.371
Interaction (Drawing × Work ethic)	0.13	0.07	1.84	.066
Interaction (Distribution × Work ethic)	0.00	0.05	-0.06	.952
Interaction (Generation × Distribution × Work ethic)	**-0.10**	**0.04**	**-2.64**	**.009**
Interaction (Drawing × Distribution × Work ethic)	-0.07	0.07	-0.92	.359
Lasting conceptual knowledge				
Prior conceptual knowledge as moderator				
Generation contrast[a]	0.06	0.04	1.47	.144
Drawing contrast[b]	0.10	0.07	1.37	.173
Distribution contrast[c]	-0.04	0.06	-0.68	.498
Prior conceptual knowledge	**0.24**	**0.06**	**4.02**	**<.001**
Interaction (Generation × Distribution)	-0.04	0.04	-0.86	.390
Interaction (Drawing × Distribution)	0.00	0.08	0.01	.996
Interaction (Generation × Prior conceptual knowledge)	0.02	0.05	0.39	.696
Interaction (Drawing × Prior conceptual knowledge)	0.04	0.07	0.48	.629

Variable	β	*SE*	*t*	*p*
Interaction (Distribution × Prior conceptual knowledge)	0.03	0.06	0.57	.566
Interaction (Generation × Distribution × Prior conceptual knowledge)	0.03	0.04	0.84	.400
Interaction (Drawing × Distribution × Prior conceptual knowledge)	0.02	0.07	0.35	.730
Academic self-concept as moderator				
Generation contrast[a]	0.06	0.04	1.42	.157
Drawing contrast[b]	0.08	0.07	1.14	.257
Distribution contrast[c]	-0.03	0.06	-0.51	.612
Academic self-concept	0.11	0.06	1.75	.082
Prior conceptual knowledge	**0.24**	**0.06**	**4.00**	**<.001**
Interaction (Generation × Distribution)	-0.04	0.04	-0.85	.398
Interaction (Drawing × Distribution)	0.00	0.08	-0.04	.970
Interaction (Generation × Academic self-concept)	-0.02	0.05	-0.39	.698
Interaction (Drawing × Academic self-concept)	0.04	0.07	0.55	.580
Interaction (Distribution × Academic self-concept)	0.05	0.07	0.72	.473
Interaction (Generation × Distribution × Academic self-concept)	-0.04	0.05	-0.98	.329
Interaction (Drawing × Distribution × Academic self-concept)	-0.04	0.07	-0.53	.593
Work ethic as moderator				
Generation contrast[a]	0.06	0.04	1.53	.127
Drawing contrast[b]	0.09	0.07	1.28	.200
Distribution contrast[c]	-0.04	0.06	-0.67	.503
Work ethic	0.07	0.06	1.06	.288
Prior conceptual knowledge	**0.25**	**0.06**	**4.07**	**<.001**
Interaction (Generation × Distribution)	-0.03	0.04	-0.75	.451
Interaction (Drawing × Distribution)	0.01	0.08	0.13	.898
Interaction (Generation × Work ethic)	0.00	0.04	0.08	.937
Interaction (Drawing × Work ethic)	0.13	0.08	1.68	.094
Interaction (Distribution × Work ethic)	0.02	0.06	0.40	.690
Interaction (Generation × Distribution × Work ethic)	-0.04	0.05	-0.86	.394
Interaction (Drawing × Distribution × Work ethic)	-0.02	0.08	-0.32	.751

Note. Significant results are highlighted in bold letters; $p < .050$.

[a] -2 = restudy, 1 = teaching-only, 1 = teaching + drawing. [b] 0 = restudy, -1 = teaching-only, 1 = teaching + drawing. [c] -1 = after-study, 1 = distributed.

Appendix F

Means, Standard Deviations, and Correlations With Confidence Intervals Regarding all Variables

	Variable	*M*	*SD*	1	2	3	4
1	Academic self-concept in physics	2.75	0.56	–			
2	Physics work ethic	2.96	0.50	**.55** [.47, .63]	–		
3	Prior conceptual knowledge	7.72	3.48	.06 [-.05, .18]	-.02 [-.13, .10]	–	
4	Prior monitoring accuracy	9.42	6.62	**.20** [.09, .31]	**.14** [.03, .25]	**-.53** [-.61, -.44]	–
5	Students' explanations completeness[a]	4.86	2.71	**.21** [.07, .34]	**.19** [.06, .32]	.00 [-.13, .14]	-.07 [-.21, .07]
6	Students' explanations elaboration[a]	3.94	5.41	.06 [-.08, .20]	**.17** [.04, .30]	.01 [-.13, .15]	-.13 [-.27, .00]
7	Students' explanations correctness[a]	88.92	14.95	-.01 [-.15, .13]	-.06 [-.20, .08]	-.02 [-.16, .11]	.04 [-.09, .18]
8	Students' drawings completeness[a]	5.90	2.69	**.22** [.03, .40]	.17 [-.02, .35]	.07 [-.13, .26]	-.16 [-.34, .03]
9	Students' drawings elaboration[a]	2.31	3.08	.01 [-.18, .20]	.02 [-.17, .21]	.10 [-.09, .29]	-.02 [-.21, .18]
10	Students' drawings correctness[a]	82.16	22.15	.14 [-.06, .32]	.18 [-.01, .36]	.04 [-.15, .23]	-.12 [-.30, .08]
11	Students' restudy notes completeness[a]	3.29	2.23	-.08 [-.27, .12]	.04 [-.16, .24]	.18 [-.02, .36]	-.13 [-.32, .07]
12	Students' restudy notes elaboration[a]	2.38	2.33	-.06 [-.25, .15]	.00 [-.20, .20]	.03 [-.17, .23]	-.10 [-.29, .10]
13	Students' restudy notes correctness[a]	92.32	12.30	.14 [-.06, .33]	.10 [-.10, .29]	-.12 [-.31, .08]	.10 [-.11, .29]
14	Immediate conceptual knowledge	15.86	5.28	**.25** [.14, .36]	**.14** [.03, .25]	**.23** [.12, .33]	-.11 [-.22, .00]
15	Immediate monitoring accuracy	-0.35	7.46	.01 [-.10, .12]	.04 [-.07, .15]	**-.20** [-.30, -.09]	**.40** [.30, .49]
16	Lasting conceptual knowledge	11.98	4.71	**.18** [.06, .30]	.10 [-.02, .22]	**.30** [.19, .41]	-.02 [-.14, .10]
17	Lasting monitoring accuracy	1.33	6.94	.06 [-.06, .18]	.06 [-.06, .18]	-.09 [-.21, .04]	**.24** [.12, .35]

	Variable	5	6	7	8	9	10
1	Academic self-concept in physics						
2	Physics work ethic						
3	Prior conceptual knowledge						
4	Prior monitoring accuracy						
5	Students' explanations completeness[a]	–					
6	Students' explanations elaboration[a]	**.28** [.15, .40]	–				
7	Students' explanations correctness[a]	.06 [-.07, .20]	.08 [-.06, .21]	–			
8	Students' drawings completeness[a]	**.39** [.21, .54]	.05 [-.14, .24]	.09 [-.10, .28]	–		
9	Students' drawings elaboration[a]	-.03 [-.22, .16]	**.31** [.13, .47]	.10 [-.09, .29]	.00 [-.19, .19]	–	
10	Students' drawings correctness[a]	**.23** [.04, .40]	-.05 [-.23, .14]	-.06 [-.24, .13]	**.43** [.27, .58]	**-.44** [-.58, -.28]	–
11	Students' restudy notes completeness[a]	–	–	–	–	–	–
12	Students' restudy notes elaboration[a]	–	–	–	–	–	–
13	Students' restudy notes correctness[a]	–	–	–	–	–	–
14	Immediate conceptual knowledge	**.27** [.14, .39]	.08 [-.05, .21]	.13 [.00, .26]	**.33** [.15, .49]	**.23** [.04, .40]	.15 [-.04, .33]
15	Immediate monitoring accuracy	**-.15** [-.28, -.02]	-.04 [-.17, .10]	-.05 [-.19, .08]	**-.30** [-.46, -.11]	-.18 [-.36, .01]	-.18 [-.36, .01]
16	Lasting conceptual knowledge	**.18** [.04, .32]	.05 [-.09, .19]	.08 [-.07, .22]	**.25** [.05, .43]	.19 [-.02, .38]	.09 [-.11, .29]
17	Lasting monitoring accuracy	.03 [-.12, -.12]	.05 [-.10, .19]	-.04 [-.18, .11]	-.14 [-.34, .07]	-.03 [-.23, .18]	-.18 [-.37, .03]

	Variable	11	12	13	14	15	16
1	Academic self-concept in physics						
2	Physics work ethic						
3	Prior conceptual knowledge						
4	Prior monitoring accuracy						
5	Students' explanations completeness[a]						
6	Students' explanations elaboration[a]						
7	Students' explanations correctness[a]						
8	Students' drawings completeness[a]						
9	Students' drawings elaboration[a]						
10	Students' drawings correctness[a]						
11	Students' restudy notes completeness[a]	–					
12	Students' restudy notes elaboration[a]	**.29** [.10, .46]	–				
13	Students' restudy notes correctness[a]	.07 [-.13, .26]	-.02 [-.22, .18]	–			
14	Immediate conceptual knowledge	.17 [-.02, .36]	.15 [-.04, .34]	.13 [-.07, .32]	–		
15	Immediate monitoring accuracy	-.15 [-.33, .05]	-.18 [-.37, .01]	-.04 [-.23, .16]	**-.62** [-.68, .55]	–	
16	Lasting conceptual knowledge	.13 [-.09, .33]	.13 [-.09, .33]	**.21** [.00, .40]	**.56** [.48, .64]	**-.29** [-.40, -.18]	–
17	Lasting monitoring accuracy	-.04 [-.25, .17]	-.15 [-.35, .07]	.03 [-.19, .24]	**-.29** [-.40, -.18]	.44 [.34, .53]	**-.49** [-.58, -.40]

Note. The data in this table are based on the raw data. Significant correlations are highlighted in bold; $p < .050$.
[a] Variables related to the learning activity (restudy, teaching-only, teaching + drawing).

8.3 Appendix Study 3

Appendix A

Overview of the Overall Dataset

Variable	Number/type of items	Used in present paper
Demographical data		
Age	1 item, rated from 10 to 17 years	×
Sex	1 item, closed answer format	×
Native language	1 item, open answer format	×
Schooltype & level	1 item, closed answer format	–
Grade in physics, last report	1 item, rated from 1 to 6	–
Graduation of parents	1 item, closed answer format	–
Students' prerequisites		
Interest in physics	4 items, rated on a 4-point Likert scale	×
Physics work ethic	4 items, rated on a 4-point Likert scale	×
Academic self-concept in physics	4 items, rated on a 4-point Likert scale	×
ICT interest	4 items, rated on a 4-point Likert scale	×
Experience with touch devices	4 items, rated on a 5-point Likert scale	–
Monitoring rating	1 item, rated from 0 to 30 points	×
Prior		
Conceptual knowledge	15 items, multiple-choice-format	×
Monitoring accuracy[a]		×
Big five personality trait domains		
Extraversion	3 items, rated on a 5-point Likert scale	–
Compatibility	3 items, rated on a 5-point Likert scale	–
Conscientiousness	3 items, rated on a 5-point Likert scale	–
Neuroticism	3 items, rated on a 5-point Likert scale	–
Openness	3 items, rated on a 5-point Likert scale	–
Perceived ratings regarding the teaching unit		
Cognitve load		
Active cognitive load	1 item, rated on a 9-point Likert scale	×
Passive cognitive load	1 item, rated on a 9-point Likert scale	×
Affect		
Arousal	1 item, rated on a 9-point Likert scale	×
Mood	1 item, rated on a 9-point Likert scale	×
Teaching quality		×
Cognitive activation	6 items, rated on a 4-point Likert scale	×
Disturbances	3 items, rated on a 4-point Likert scale	×
Teacher monitoring	5 items, rated on a 4-point Likert scale	×
Teacher support	4 items, rated on a 4-point Likert scale	×
Characteristics of students' explanations		
Completeness	1 coded item, range 0 to 10 points	×
Elaboration	1 coded item, 1 point per elaboration	×
Correctness	1 coded item, percentage value	×
Characteristics of students' drawings		
Completeness	1 coded item, range 0 to 10 points	×
Elaboration	1 coded item, 1 point per elaboration	×
Correctness	1 coded item, percentage value	×
Characteristics of students' restudy notes regarding the generative learning activity phase		
Completeness	1 coded item, range 0 to 10 points	×
Elaboration	1 coded item, 1 point per elaboration	×
Correctness	1 coded item, percentage value	×
Retrieval quiz		

Variable	Number/type of items	Used in present paper
Retrieval success	1 coded item, range 0 to 13 retrieved core concepts	×
Characteristics of students' restudy notes regarding the retrieval practice phase		
Completeness	1 coded item, range 0 to 10 points	×
Elaboration	1 coded item, 1 point per elaboration	×
Correctness	1 coded item, percentage value	×
Perceived ratings immediately after interventions		
Active cognitive load	1 item, rated on a 9-point Likert scale	×
Passive cognitive load	1 item, rated on a 9-point Likert scale	×
Task interest	2 items, rated on a 4-point Likert scale	−
Task enjoyment	2 items, rated on a 4-point Likert scale	−
Arousal	1 item, rated on a 9-point Likert scale	−
Mood	1 item, rated on a 9-point Likert scale	−
Monitoring rating	1 item, rated from 0 to 30 points	×
Immediate		
Conceptual knowledge	15 items, multiple-choice-format	×
Monitoring accuracy[a]		×
Perceived rating delayed after learning activity		
Monitoring rating	1 item, rated from 0 to 30 points	×
Lasting		
Conceptual knowledge	15 items, multiple choice format	×
Monitoring accuracy[a]		×

Note. All variables are presented in chronological order. [a]Monitoring accuracy is determined by the difference between students' estimated and actual performance.

Appendix B

Summary of Means and Standard Deviations for All Measurements by Condition

Variable	Non-Generation				Generation			
	Non-Retrieval		Retrieval		Non-Retrieval		Retrieval	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Students' prerequisites								
Interest in physics (1-4)	2.65	0.69	2.74	0.60	2.66	0.58	2.77	0.70
Physics work ethic (1-4)	2.71	0.55	2.82	0.53	2.77	0.59	2.93	0.59
Academic self-concept in physics (1-4)	2.47	0.74	2.58	0.62	2.54	0.72	2.73	2.75
ICT interest (1-4)	3.06	0.56	3.24	0.49	3.13	0.56	3.07	0.53
Prior								
Conceptual knowledge (0-30)	8.70	3.59	9.10	4.28	8.78	3.53	9.04	3.89
Monitoring accuracy[a] (-30-30)	6.31	7.43	7.10	6.53	7.24	6.65	8.07	7.48
Perceived ratings regarding the teaching unit								
Cognitive load								
Active cognitive load (1-9)	5.61	1.86	5.78	1.76	5.71	1.84	6.02	1.81
Passive cognitive load (1-9)	3.52	1.98	3.34	1.73	3.33	1.74	3.29	1.70
Affect								
Arousal (1-9)	4.47	2.17	4.97	1.87	4.47	2.08	4.88	2.02
Mood (1-9)	6.06	2.00	6.07	2.01	6.15	1.83	6.12	1.96
Teaching quality								
Cognitive activation (1-4)	2.61	0.52	2.62	0.53	2.55	0.57	2.72	0.49
Disturbances (1-4)	1.80	0.74	1.84	0.64	1.87	0.69	1.97	0.71
Teacher monitoring (1-4)	2.94	0.55	2.89	0.62	2.85	0.59	2.85	0.59
Teacher support (1-4)	3.27	0.58	3.17	0.67	3.07	0.64	3.14	0.62
Characteristics of students' explanations								
Completeness (0-10 points)	–	–	–	–	4.49	2.67	4.86	2.62
Elaboration (each 1 point)	–	–	–	–	3.91	3.96	5.14	5.51
Correctness (percentage)	–	–	–	–	91.23	14.54	90.79	15.16
Characteristics of students' drawings								
Completeness (0-10 points)	–	–	–	–	6.69	2.43	6.52	2.59
Elaboration (each 1 point)	–	–	–	–	0.74	1.17	1.05	2.18

Variable	Non-Generation				Generation			
	Non-Retrieval		Retrieval		Non-Retrieval		Retrieval	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Correctness (percentage)	–	–	–	–	91.50	9.46	90.85	11.51
Characteristics of students' restudy notes regarding the generative learning activity phase								
Completeness (0-10 points)	3.72	2.24	4.13	2.18	–	–	–	–
Elaboration (each 1 point)	1.16	1.42	1.24	1.28	–	–	–	–
Correctness (percentage)	89.98	19.73	93.00	12.70	–	–	–	–
Retrieval quiz								
Retrieval success (0-12 retrieved concepts)			5.54	0.90			5.64	0.92
Characteristics of students' restudy notes regarding the retrieval practice phase								
Completeness (0-10 points)	3.79	2.35	–	–	3.95	2.13	–	–
Elaboration (each 1 point)	1.14	1.51	–	–	0.76	0.99	–	–
Correctness (percentage)	91.76	15.69	–	–	93.48	13.07	–	–
Perceived ratings regarding the interventions								
Cognitive load								
Active cognitive load (1-9)	5.12	2.40	5.86	2.06	5.70	2.09	6.10	2.05
Passive cognitive load (1-9)	3.35	2.02	4.21	1.94	4.24	1.76	4.53	2.09
Immediate								
Conceptual knowledge (0-30)	15.20	5.26	15.94	5.03	15.53	4.76	16.89	5.64
Monitoring accuracy[a] (-30-30)	-0.81	7.32	-1.12	6.83	-0.56	6.79	-1.43	7.79
Lasting								
Conceptual knowledge (0-30)	12.64	4.32	13.26	5.11	12.47	4.84	13.49	4.89
Monitoring accuracy[a] (-30-30)	-0.78	6.55	-0.10	6.41	1.15	6.97	0.88	7.23

Note. [a]Monitoring accuracy reflects the difference between students' estimated and actual performance. Negative values indicate underestimation, positive values indicate overestimation, and a value of zero represents perfect judgment accuracy.

Appendix C

Correlations Separated by Experimental Condition

	Variable	1	2	3	4	5
1	Prior conceptual knowledge	–				
2	Prior monitoring accuracy	(-.40*|-.53*|-.47*|-.69*)	–			
3	Students' explanations completeness[a]	(-|-|.12|-.01)	(-|-|-.08|.06)	–		
4	Students' explanations elaboration[a]	(-|-|-.19|-.03)	(-|-|.10|.05)	(-|-|.34*|.37*)	–	
5	Students' explanations correctness[a]	(-|-|.11|-.16)	(-|-|-.07|.12)	(-|-|.21|.23*)	(-|-|.12|.08)	–
6	Students' drawings completeness[a]	(-|-|-.10|-.05)	(-|-|.18|.13)	(-|-|.28*|.31*)	(-|-|.10|.09)	(-|-|.22|.11)
7	Students' drawings elaboration[a]	(-|-|.06|-.07)	(-|-|.01|-.05)	(-|-|-.03|.01)	(-|-|.07|.27*)	(-|-|-.02|.17)
8	Students' drawings correctness[a]	(-|-|.04|-.24*)	(-|-|-.27*|.10)	(-|-|.30*|.35*)	(-|-|.14|.23*)	(-|-|.14|.18)
9	Students‘ restudy notes completeness[a]	(.17|.02|-|-)	(-.08|-.11|-|-)	(-|-|-|-)	(-|-|-|-)	(-|-|-|-)
10	Students‘ restudy notes elaboration[a]	(.16|.06|-|-)	(.05|.04|-|-)	(-|-|-|-)	(-|-|-|-)	(-|-|-|-)
11	Students‘ restudy notes correctness[a]	(.12|.13|-|-)	(-.16|-.11|-|-)	(-|-|-|-)	(-|-|-|-)	(-|-|-|-)
12	Retrieval success[b]	(-|.18|-|.01)	(-|-.05|-|.02)	(-|-|-|.28*)	(-|-|-|.09)	(-|-|-|.06)
13	Students‘ restudy notes completeness[b]	(.11|-|-.03|-)	(-.08|-|-.18|-)	(-|-|.49*|-)	(-|-|.19|-)	(-|-|.17|-)
14	Students‘ restudy notes elaboration[b]	(.13|-|.01|-)	(.10|-|-.16|-)	(-|-|.29*|-)	(-|-|.42*|-)	(-|-|.24*|-)
15	Students‘ restudy notes correctness[b]	(.08|-|.11|-)	(-.04|-|-.06|-)	(-|-|.09|-)	(-|-|-.06|-)	(-|-|.43*|-)
16	Active cognitive load[c]	(-.04|-.08|.05|-.31*)	(-.14|.10|-.05|.35*)	(-|-|.08|.24*)	(-|-|.15|.16)	(-|-|.02|.03)
17	Passive cognitive load[c]	(-.28*|-.25*|.00|-.09)	(.22|.18|-.06|-.01)	(-|-|-.01|.15)	(-|-|.26*|-.02)	(-|-|-.09|-.12)
18	Immediate conceptual knowledge	(.11|.44*|.52*|.27*)	(.06|-.17|-.21|-.08)	(-|-|.27*|.27*)	(-|-|.04|.30*)	(-|-|.14|.06)
19	Immediate monitoring accuracy	(.06|-.14|-.29|-.25*)	(.43*|.43*|.42*|.52*)	(-|-|-.15|-.01)	(-|-|.12|-.03)	(-|-|-.18|-.03)
20	Lasting conceptual knowledge	(.22|.48*|.31|.35*)	(-.05|-.33*|-.13|-.23*)	(-|-|.34*|.12)	(-|-|.24*|.11)	(-|-|.23|-.06)
21	Lasting monitoring accuracy	(.07|-.24|-.12|-.20)	(.36*|.48*|.32*|.33*)	(-|-|-.26*|.17)	(-|-|-.10|.08)	(-|-|-.26*|.10)

	Variable	6	7	8	9	10
1	Prior conceptual knowledge					
2	Prior monitoring accuracy					
3	Students' explanations completeness[a]					
4	Students' explanations elaboration[a]					
5	Students' explanations correctness[a]					
6	Students' drawings completeness[a]	–				
7	Students' drawings elaboration[a]	(-|-|.05|-.10)	–			
8	Students' drawings correctness[a]	(-|-|.30*|.41*)	(-|-|.06|.13)	–		
9	Students' restudy notes completeness[a]	(-|-|-|-)	(-|-|-|-)	(-|-|-|-)	–	
10	Students' restudy notes elaboration[a]	(-|-|-|-)	(-|-|-|-)	(-|-|-|-)	(.39*|.24*|-|-)	–
11	Students' restudy notes correctness[a]	(-|-|-|-)	(-|-|-|-)	(-|-|-|-)	(.44*|.48*|-|-)	(.27*|.21|-|-)
12	Retrieval success[b]	(-|-|-|.21*)	(-|-|-|-.04)	(-|-|-|.18)	(-|.34*|-|-)	(-|.46*|-|-)
13	Students' restudy notes completeness[b]	(-|-|.25*|-)	(-|-|.05|-)	(-|-|.34*|-)	(.77*|-|-|-)	(.49*|-|-|-)
14	Students' restudy notes elaboration[b]	(-|-|.07|-)	(-|-|-.15|-)	(-|-|.23*|-)	(.36*|-|-|-)	(.80*|-|-|-)
15	Students' restudy notes correctness[b]	(-|-|-.20|-)	(-|-|.19|-)	(-|-|.05|-)	(.34*|-|-|-)	(.18|-|-|-)
16	Active cognitive load[c]	(-|-|-.03|.08)	(-|-|.13|.03)	(-|-|.10|.11)	(.09|.06|-|-)	(.21|-.05|-|-)
17	Passive cognitive load[c]	(-|-|-.12|-.04)	(-|-|-.08|.10)	(-|-|.08|.02)	(-.15|.16|-|-)	(.01|-.08|-|-)
18	Immediate conceptual knowledge	(-|-|.20|.11)	(-|-|.27*|.23*)	(-|-|.28*|.30*)	(.41*|.32*|-|-)	(.50*|.35*|-|-)
19	Immediate monitoring accuracy	(-|-|-.07|.09)	(-|-|.10|-.20)	(-|-|-.34*|-.13)	(-.13|-.12|-|-)	(-.15|-.11|-|-)
20	Lasting conceptual knowledge	(-|-|.28*|.07)	(-|-|.34*|.12)	(-|-|.20|.06)	(.33*|.34*|-|-)	(.48*|.19|-|-)
21	Lasting monitoring accuracy	(-|-|-.09|.03)	(-|-|.00|-.11)	(-|-|-.13|.07)	(-.17|-.22|-|-)	(-.04|.05|-|-)

	Variable	11	12	13	14	15
1	Prior conceptual knowledge					
2	Prior monitoring accuracy					
3	Students' explanations completeness[a]					
4	Students' explanations elaboration[a]					
5	Students' explanations correctness[a]					
6	Students' drawings completeness[a]					
7	Students' drawings elaboration[a]					
8	Students' drawings correctness[a]					
9	Students' restudy notes completeness[a]					
10	Students' restudy notes elaboration[a]					
11	Students' restudy notes correctness[a]	–				
12	Retrieval success[b]	(-|.30*|-|-)	–			
13	Students' restudy notes completeness[b]	(.37*|-|-|-)	(-|-|-|-)	–		
14	Students' restudy notes elaboration[b]	(.24*|-|-|-)	(-|-|-|-)	(.42*|-|.39*|-)	–	
15	Students' restudy notes correctness[b]	(.67*|-|-|-)	(-|-|-|-)	(.35*|-|.20|-)	(.22|-|.17|-)	–
16	Active cognitive load[c]	(.11|.01|-|-)	(-|-.05|-|.30*)	(.04|-|-.01|-)	(.15|-|.19|-)	(-.05|-|.06|-)
17	Passive cognitive load[c]	(-.22*|.01|-|-)	(-|.05|-|.18)	(-.17|-|-.01|-)	(.04|-|.22*|-)	(-.14|-|-.03|-)
18	Immediate conceptual knowledge	(.25*|.17|-|-)	(-|.50*|-|.37*)	(.47*|-|.10|-)	(.44*|-|.14|-)	(.17|-|.05|-)
19	Immediate monitoring accuracy	(-.08|-.01|-|-)	(-|-.12|-|-.14)	(-.17|-|-.28*|-)	(-.08|-|-.22*|-)	(-.07|-|-.11|-)
20	Lasting conceptual knowledge	(.15|.30*|-|-)	(-|.39*|-|.21)	(.43*|-|.10|-)	(.41*|-|.12|-)	(.13|-|-.10|-)
21	Lasting monitoring accuracy	(-.10|-.23|-|-)	(-|-.13|-|.00)	(-.21|-|-.29*|-)	(.06|-|-.27*|-)	(-.09|-|.09|-)

	Variable	16	17	18	19	20
1	Prior conceptual knowledge					
2	Prior monitoring accuracy					
3	Students' explanations completeness[a]					
4	Students' explanations elaboration[a]					
5	Students' explanations correctness[a]					
6	Students' drawings completeness[a]					
7	Students' drawings elaboration[a]					
8	Students' drawings correctness[a]					
9	Students' restudy notes completeness[a]					
10	Students' restudy notes elaboration[a]					
11	Students' restudy notes correctness[a]					
12	Retrieval success[b]					
13	Students' restudy notes completeness[b]					
14	Students' restudy notes elaboration[b]					
15	Students' restudy notes correctness[b]					
16	Active cognitive load[c]	–				
17	Passive cognitive load[c]	(.20|.24*|.13|.36*)	–			
18	Immediate conceptual knowledge	(.04|-.19|.30*|.09)	(-.08|.03|.09|.03)	–		
19	Immediate monitoring accuracy	(-.11|.34*|.11|.20)	(.11|-.05|-.05|-.08)	(-.48*|-.48*|-.40*|-.53*)	–	
20	Lasting conceptual knowledge	(.11|-.04|.25*|-.06)	(.02|.01|-.05|-.03)	(.66*|.52*|.56*|.67*)	(-.24*|-.19|-.08|-.34*)	–
21	Lasting monitoring accuracy	(-.14|.08|-.21|.11)	(.02|-.13|.10|-.14)	(-.15|-.12|-.10|-.26*)	(.51*|.35*|.23*|.38*)	(-.46*|-.56*|-.50*|-.54*)

Note. The data in this table are based on the raw data. Numbers in brackets represent the correlations separated for experimental conditions: first = non-generation + non-retrieval; second = non-generation + retrieval; third = generation + non-retrieval; fourth = generation + retrieval.

[a] Variables regarding the generative learning activity phase (non-generation, generation).

[b] Variables regarding the retrieval practice phase (non-retrieval, retrieval).

[c] Variables regarding the interventions (non-generation + non-retrieval, non-generation + retrieval, generation + non-retrieval, generation + retrieval).

$^{*}p < .050$.

Appendix D

Structural Equation Modeling Results for Conceptual Knowledge and Monitoring Accuracy

Variable	β	*SE*	*p*
Conceptual knowledge			
Immediate			
Generation[a]	0.05	0.21	.797
Retrieval[b]	0.11	0.11	.307
Interaction (Generation × Retrieval)	0.12	0.20	.537
Prior conceptual knowledge	**0.33**	**0.08**	**<.001**
Lasting			
Generation[a]	-0.06	0.17	.722
Retrieval[b]	0.06	0.23	.783
Interaction (Generation × Retrieval)	0.12	0.24	.598
Prior conceptual knowledge	**0.35**	**0.08**	**<.001**
Monitoring accuracy			
Immediate			
Generation[a]	-0.01	0.13	.969
Retrieval[b]	-0.08	0.25	.748
Interaction (Generation × Retrieval)	-0.09	0.25	.726
Prior monitoring accuracy	**0.46**	**0.04**	**<.001**
Lasting			
Generation[a]	0.25	0.15	.093
Retrieval[b]	0.08	0.13	.551
Interaction (Generation × Retrieval)	-0.18	0.16	.268
Prior monitoring accuracy	**0.37**	**0.08**	**<.001**

Note. Significant results are highlighted in bold letters; $p < .050$.

[a] 0 = non-generative, 1 = generative. [b] 0 = non-retrieval, 1 = retrieval.

Appendix E

Summary of the Moderation Analyses With Characteristics of Students' Explanations and Restudy Notes Regarding Immediate and Lasting Conceptual Knowledge

Variable	β	*SE*	*p*
Immediate conceptual knowledge			
Completeness as moderator			
Generation[a]	0.07	0.10	.465
Retrieval[b]	**0.14**	**0.04**	**.001**
Completeness	**0.27**	**0.07**	**<.001**
Interaction (Generation × Retrieval × Completeness)	-0.07	0.07	.316
Prior conceptual knowledge	**0.31**	**0.08**	**<.001**
Elaboration as moderator			
Generation[a]	-0.07	0.17	.678
Retrieval[b]	**0.14**	**0.06**	**.012**
Elaboration	0.25	0.13	.053
Interaction (Generation × Retrieval × Elaboration)	-0.04	0.12	.738
Prior conceptual knowledge	**0.34**	**0.08**	**<.001**
Correctness as moderator			
Generation[a]	0.13	0.11	.233
Retrieval[b]	**0.17**	**0.05**	**<.001**
Correctness	0.09	0.07	.208
Interaction (Generation × Retrieval × Correctness)	-0.11	0.09	.201
Prior conceptual knowledge	**0.32**	**0.08**	**<.001**
Lasting conceptual knowledge			
Completeness as a moderator			
Generation[a]	-0.03	0.13	.836
Retrieval[b]	0.11	0.14	.438
Completeness	**0.31**	**0.08**	**<.001**
Interaction (Generation × Retrieval × Completeness)	**-0.20**	**0.06**	**.002**
Prior conceptual knowledge	**0.33**	**0.07**	**<.001**
Elaboration as moderator			
Generation[a]	-0.22	0.15	.134
Retrieval[b]	0.14	0.15	.321
Elaboration	**0.38**	**0.12**	**.002**
Interaction (Generation × Retrieval × Elaboration)	**-0.28**	**0.10**	**.007**
Prior conceptual knowledge	**0.36**	**0.08**	**<.001**
Correctness as moderator			
Generation[a]	0.03	0.12	.796
Retrieval[b]	0.12	0.12	.355
Correctness	**0.16**	**0.06**	**.004**
Interaction (Generation × Retrieval × Correctness)	**-0.18**	**0.07**	**.006**
Prior conceptual knowledge	**0.33**	**0.08**	**<.001**

Note. Significant results are highlighted in bold letters; $p < .050$.

[a] 0 = non-generative, 1 = generative. [b] 0 = non-retrieval, 1 = retrieval.