

Beyond the Screen:
**How Feedback-Related, Learner-Related, and Context-Related Factors
Determine the Effectiveness of Computer-Based Feedback on Learning**

Dissertation
zur Erlangung des Doktorgrades
der Wirtschafts- und Sozialwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen

vorgelegt von
Salome Wagner

Tübingen
2024

1. Betreuer:	Prof. Dr. Andreas Lachner
2. Betreuerin:	Prof.in Dr.in Katharina Scheiter
Tag der mündlichen Prüfung:	29.02.2024
Dekan:	Prof. Dr. Ansgar Thiel
1. Gutachter:	Prof. Dr. Andreas Lachner
2. Gutachterin:	Prof.in Dr.in Taiga Brahm

Danksagung

Ich danke all den vielen Menschen, die mich während der letzten vier Jahre begleitet und unterstützt haben.

Ein ganz besonders großer Dank geht an Prof. Dr. Andreas Lachner, meinen Doktorvater, der diesem Begriff alle Ehre macht. Danke, Andi, für dein stets konstruktives Feedback, deine Geduld, Motivation und Fürsorge, für deine persönliche Art und dass du mein Kaffee-Verbündeter bist. Danke, dass du mich gefordert und gefördert hast und für deine fabelhafte und enge Betreuung, die ich mir nicht besser hätte wünschen können.

An dieser Stelle danke ich auch ganz herzlich Prof. Dr. Katharina Scheiter, meiner Zweitbetreuerin. Danke für deinen Zuspruch, deine Ermutigung, deine herzliche Art und dein sehr hilfreiches, durchdachtes Feedback, durch das meine Dissertation wesentlich besser wurde.

Ein großer Dank geht auch an Prof. Dr. Taiga Brahm. Danke, dass du mein zweites Gutachten übernimmst, für die gemeinsamen Mittagessen, dein Feedback und deine Ermunterung in den letzten Jahren und den stets interessanten und erfreulichen Austausch mit dir.

Danke auch an Prof. Dr. Thorsten Bohl, dass du den Prüfungsvorsitz für meine Disputation trotz deines vollen Terminkalenders übernimmst. Ich schätze deine stets inspirierenden Gedanken und dein konstruktives Feedback sehr und blicke mit Freude auf unsere gemeinsame Zeit am IfE und an der TÜSE zurück.

Große Dankbarkeit empfinde ich auch für die Community von LEAD, der TüSE, der Abteilung Schulpädagogik des IfE und der EARLI. Ich habe hier so viele tolle Menschen kennengelernt, von denen ich viel Unterstützung erfahren habe, mit denen ich viel Spaß hatte und die mir auch gezeigt haben, warum die Wissenschaft so wunderbar ist und sich die ganze Arbeit lohnt. Ein großes Danke an alle, deren Namen ich hier leider nicht alle aufführen kann.

Ebenso danke ich der AG-Digitalisierung, meiner Arbeitsfamilie, mit der ich durch gute und schlechte Zeiten gegangen bin und wegen der ich immer gerne ins Büro komme. Danke für euren Support, eure Geduld mit mir und die schöne gemeinsame Zeit.

Ganz besonders danke ich an dieser Stelle der anderen Salome, Franzi und Armin, meinen Freunden und Weggefährten, ohne die diese Promotionszeit wesentlich härter gewesen wäre.

Danksagung

Danke für eure Ratschläge, euer Feedback, eure Tipps und Tricks, euren Beistand, eure Wertschätzung und unsere gemeinsame Zeit. Danke, dass ihr mich immer aufgemuntert habt und für mich da wart, wenn es mir mal nicht gut ging, dass ich mich bei euch ausheulen durfte und ihr mit mir gelitten, mitgefiebert und gefeiert habt. So gute Freunde wie euch findet man selten. Ich bin von Herzen dankbar euch zu haben.

In tiefer Verbundenheit danke ich Dr. Leonie Sibley und Dr. Iris Backfisch, meinen daily supervisors und Büromitbewohnerinnen. Danke, dass ihr mich von Anfang an so lieb aufgenommen habt, als es nur wir drei bei Andi waren, für eure ganze Zeit, dass ich mit jeder dummen Frage zu euch kommen konnte, für eure R-Skripte und Vorlagen, dass ihr diese Dissertation in sämtlichen Fassungen gelesen habt und für euer stets hilfreiches Feedback. Danke, dass ihr mich anfeuert, wie keine anderen, für eure Motivation und aufbauenden Worte, für eure fabelhafte Gesellschaft im Büro, für eure Fürsorge und Unterstützung in allen Lebenslagen und vor allem für eure Freundschaft. Ohne euch hätte ich es nie so weit geschafft. Ich bin unendlich dankbar, dass ich euch durch diese Promotion kennenlernen durfte.

Ein riesiges Dankeschön geht an meine Familie, allen voraus meine Eltern, die mir das alles ermöglicht haben. Danke für eure Unterstützung in jeglicher Hinsicht, euer Interesse an meiner Arbeit, dass ihr wirklich zu jeder Tages- und Nachtzeit für mich da seid, euren unendlichen Glauben an mich und eure Liebe. Ebenso danke ich meinen Brüdern, dass ihr immer hinter mir steht, für mich da seid und mich umsorgt, dass ihr mein Gerede über die Dissertation vier Jahre lang ertragen habt und ich euch immer anrufen oder mich zu euch flüchten konnte, wenn es mal hart war. Danke auch an meine Oma Ella, für deine Unterstützung, deinen Beistand und deine Ermutigungen und dass du für mich versuchst, diese Welt der Wissenschaft zu verstehen.

Ich kann nicht in Worte fassen, wie sehr ich Patrick Böß danke für all die Unterstützung von Anfang an. Danke, dass du extra für mich nach Tübingen gezogen bist und dich mit mir in dieses Abenteuer gestürzt hast. Danke für deine Motivationsreden und für all die Kaffees, die du mir gemacht hast, damit ich morgens aufstehen oder die Nächte durchpowern konnte. Danke fürs Zuhören, wenn ich dich schon wieder zutexte und dir jeden Tag von meiner Promotion erzähle. Danke fürs Feiern jedes kleinen Erfolgs und für deinen Beistand bei jedem Rückschlag. Vor allem DANKE für deine große Geduld, deinen Glauben an mich, dass du mir dabei hilfst, mir meine Träume zu erfüllen und für deine Liebe.

Summary

Formative feedback is considered an important and beneficial instructional strategy for fostering students' learning (Hattie & Timperley, 2007; Shute, 2008; Wisniewski et al., 2020). Since providing individualized, immediate feedback is time-consuming and difficult to realize, especially in heterogeneous classes or individual learning settings (Lipnevich & Panadero, 2021; Shute, 2008), educational technologies have been discussed as a useful supplement to provide computer-based feedback (Narciss, 2013). However, there are contradictory findings regarding the effectiveness of computer-based feedback (e.g., Mertens et al., 2022). Previous research suggests that the moderating impacts of various factors on the effect of computer-based feedback can explain these inconsistent findings (e.g., Kluger & DeNisi, 1996; Kuklick & Lindner, 2023; Mertens et al., 2022; Narciss, 2008; Panadero & Lipnevich, 2022; Patchan et al., 2016; Strobl et al., 2019). Initial evidence suggests that various feedback-related (e.g., Lachner et al., 2017a; Mertens et al., 2022), learner-related (e.g., Kuklick & Lindner, 2023; Narciss, 2008, 2013), and context-related (e.g., Fyfe & Rittle-Johnson, 2016; Shirah & Sidney, 2023) factors might affect the effectiveness of computer-based feedback.

These influencing factors, however, have not been systematically investigated. Since previous research focused only on investigating individual feedback-related, learner-related, or context-related potential influencing factors, it was still unclear which factors significantly determine the effectiveness of computer-based feedback on learning. Answering this question requires, on the one hand, theoretical conceptualization and, on the other hand, empirical investigation of the factors of the feedback message itself, of learners, and of the orchestrating context. These factors, which are embedded in the learning process, were investigated, both separately and in relation to each other (Narciss & Huth, 2004; Shute, 2008; Strobl et al., 2019). In order to fulfill these requirements, this dissertation adopted a multidimensional perspective. First, important feedback-related (i.e., specificity of feedback,

Summary

representation of feedback, level of elaboration, level of complexity), learner-related (i.e., prior knowledge, cognitive, metacognitive, and affective-motivational processes), and context-related (i.e., orchestration with prior instruction, sequence) potential influencing factors were identified based on the previous literature and the research findings. Second, a theoretical model was developed that encompasses the learning process using computer-based feedback and that was bordered by these potential influencing factors. Third, these factors were empirically investigated in different subject domains (i.e., writing and physics) and using different methodological approaches (theoretical conception, literature overview, meta-analysis, experimental studies).

In total, this dissertation comprises three articles for publication in international journals. The first is a theoretical paper in which the potential of computer-based feedback tools to enhance learners' expository writing quality was explored and potential feedback-related influencing factors have been identified. The second article is an empirical paper encompassing a pre-registered meta-analysis to examine the overall effectiveness of computer-based feedback tools in supporting students in improving their expository writing. Furthermore, the impact of the potential feedback-related factors identified in Paper 1 was examined as well as the potential impact of prior knowledge as a learner-related influencing factor. The third article focused on empirically investigating the impact of further potential learner-related and feedback-related influencing factors as well as on context-related factors related to students' learning in the domain of physics. Therefore, three pre-registered experiments were conducted that included manipulation of the orchestration of computer-based feedback with strategy instruction. Experiment 1 examined corrective feedback; Experiments 2 and 3 focused on elaborated feedback. In addition, Experiment 3 investigated the potential effect of the sequence in which computer-based feedback and strategy instruction were provided. In summary, through this multidimensional perspective investigating the effectiveness of computer-based feedback and how it is affected by

feedback-, learner-, and context-related factors, it was possible to adapt the new theoretical model adequately and provide theoretical implications and suggestions for future research accordingly. In addition, based on new evidence, this dissertation was able to provide recommendations for practice regarding what to consider when implementing and orchestrating computer-based feedback to foster learning effectively.

Zusammenfassung

Formatives Feedback wird als eine wichtige und förderliche instruktionale Strategie angesehen, um Lernen zu fördern (Hattie & Timperley, 2007; Shute, 2008; Wisniewski et al., 2020). Da individualisiertes, unmittelbares Feedback bereitzustellen zeitaufwändig und vor allem in heterogenen Klassen oder individuellen Lernsituationen schwierig umzusetzen ist (Lipnevich & Panadero, 2021; Shute, 2008), werden Bildungstechnologien als hilfreiche Ergänzung diskutiert, um computerbasiertes Feedback bereitzustellen (Narciss, 2013). Allerdings gibt es widersprüchliche Befunde hinsichtlich der Effektivität computerbasierten Feedbacks (siehe z. B. Mertens et al., 2022). Bisherige Forschung legt nahe, dass der moderierende Einfluss verschiedener Faktoren auf den Effekt von computerbasiertem Feedback diese inkonsistenten Befunde erklären könnte (z. B. Kluger & DeNisi, 1996; Kuklick & Lindner, 2023; Mertens et al., 2022; Narciss, 2008; Panadero & Lipnevich, 2022; Patchan et al., 2016; Strobl et al., 2019). Es gibt erste evidente Hinweise, die andeuten, dass sowohl verschiedene feedbackbezogene (z. B. Lachner et al., 2017a; Mertens et al., 2022), lernerbezogene (z. B. Kuklick & Lindner, 2023; Narciss, 2008, 2013) als auch kontextbezogene (z. B. Fyfe & Rittle-Johnson, 2016; Shirah & Sidney, 2023) Faktoren die Effektivität von computerbasiertem Feedback bedingen könnten.

Diese Einflussfaktoren wurden jedoch bisher nicht systematisch erforscht. Bisherige Forschung fokussierte sich meist auf einzelne potentielle feedback-, lerner- oder kontextbezogene Einflussfaktoren. Deshalb ist nach wie vor unklar, welche genau die Faktoren sind, die die Effektivität von computerbasiertem Feedback zur Förderung des Lernens entscheidend determinieren. Die Beantwortung dieser Frage erforderte zum einen eine theoretische Konzeptualisierung und zum anderen eine empirische Untersuchung von Faktoren der Feedbacknachricht selbst, von Lernenden und des orchestrierenden Kontexts. Dabei wurden die Faktoren, eingebettet in den Lernprozess, sowohl separat als auch deren Beziehung zueinander betrachtet (Narciss & Huth, 2004; Shute, 2008; Strobl et al., 2019).

Um diese Anforderungen zu erfüllen, wurde in dieser Dissertation eine multidimensionale Perspektive eingenommen. Zunächst wurden basierend auf bisheriger Literatur und bisherigen Forschungserkenntnissen wichtige potentielle feedbackbezogene (Spezifität, Darstellung, Elaborationslevel und Komplexitätslevel), lernerbezogene (Vorwissen, kognitive, meta-kognitive und affektiv-motivationale Prozesse) und kontextbezogene (Orchestrierung und Reihenfolge) Einflussfaktoren identifiziert. Des Weiteren wurde ein theoretisches Modell entwickelt, das den Lernprozess mittels computergestützten Feedbacks umfasst und von den herausgearbeiteten potentiellen Einflussfaktoren begrenzt wird. In einem dritten Schritt wurden diese Faktoren in den Fachdomänen Schreiben und Physik anhand unterschiedlicher methodischer Ansätze (theoretische Konzeption, Literaturüberblick, Metaanalyse, experimentelle Studien) untersucht.

Insgesamt beinhaltet diese Dissertation drei Artikel zur Veröffentlichung in internationalen Fachzeitschriften. Der erste Artikel ist ein theoretischer Artikel, in dem das Potenzial computerbasierter Feedbacktools zur Förderung des erklärenden Schreibens und mögliche feedbackbezogene Einflussfaktoren herausgearbeitet wurden. Der zweite Artikel ist eine präregistrierte Metaanalyse anhand derer die grundsätzliche Effektivität von computerbasiertem Feedback zur Verbesserung von Schreibkompetenzen statistisch untersucht wurde. Zusätzlich wurde im Rahmen dieser Metaanalyse der Einfluss der in Artikel 1 identifizierten möglichen feedbackbezogenen Einflussfaktoren und der potentielle Einfluss des Vorwissens als lernerbezogener Faktor auf die Effektivität von computerbasiertem Feedback zur Schreibförderung empirisch erforscht. Der dritte Artikel fokussierte die empirische Untersuchung des Einflusses weiterer möglicher lerner- und kontextbezogener Faktoren sowie eines weiteren feedbackbezogenen Faktors in der Fachdomäne Physik. Hierfür wurden drei präregistrierte Experimente durchgeführt, in denen die Orchestrierung von computerbasiertem Feedback mit Strategieinstruktion manipuliert wurde. In Experiment 1 wurde dabei korrekatives und in den Experimenten 2 und 3

Zusammenfassung

elaboriertes Feedback implementiert. In Experiment 3 wurde zusätzlich noch ein potenzieller Reihenfolgeeffekt getestet (bezogen auf die Reihenfolge, in der computerbasiertes Feedback und Strategieinstruktion den Lernenden im Lernprozess bereitgestellt werden). Zusammenfassend ermöglichte die multiperspektivische Erforschung der Effektivität computerbasierten Feedbacks und wie sich diese durch feedback-, lerner-, und kontextbezogene Faktoren verändert, das neue theoretische Modell entsprechend adäquat anzupassen und sowohl theoretische Implikationen als auch Anregungen für zukünftige Forschung abzuleiten. Des Weiteren konnten basierend auf neuer Evidenz Empfehlungen für die Praxis gegeben werden, was beim Einsatz und der Orchestrierung von computerbasiertem Feedback zu beachten ist, um Lernen effektiv zu fördern.

Table of Contents

Danksagung	III
Summary.....	V
Zusammenfassung	VIII
List of Abbreviations	XV
1 GENERAL INTRODUCTION AND THEORETICAL BACKGROUND	1
1.1 Learning.....	3
1.1.1 Types of Knowledge.....	4
1.1.2 Acquisition, Construction, and Integration of Knowledge.....	6
1.2 Feedback	9
1.2.1 Feedback Models.....	11
1.2.2 Definition of Feedback.....	15
1.2.3 Empirical Evidence of Feedback.....	16
1.3 Computer-Based Feedback	18
1.3.1 Empirical Evidence Regarding the Implementation of Computer-Based Feedback	20
1.3.2 Empirical Evidence Regarding the Orchestration of Computer-Based Feedback	22
1.3.3 Potential Influencing Factors on the Effectiveness of Computer-Based Feedback	24
1.3.3.1. Do Feedback-Related Factors make the Difference?.....	26
1.3.3.2. Do Learner-Related Factors make the Difference?	29
1.3.3.3. Do Context-Related Factors make the Difference?	31
1.4 The CAFCALI Model.....	32
1.5 Goal and Research Questions	39
1.6 Overview of the Studies of this Dissertation	41
2 FEEDBACK – JA, KLAR?!	46
2.1 Einleitung.....	48
2.2 Die Entwicklung von Schreibkompetenzen.....	49
2.3 Feedback zur Unterstützung von Revisionsprozessen.....	51
2.4 Computerbasiertes Feedback mit digitalen Medien – eine Methode zur Förderung der Schreibkompetenz von Lernenden	52
2.5 Dimensionen computerbasierter Feedbacks.....	55
2.5.1 Dimension der Textqualität	55
2.5.2 Dimension der Repräsentation.....	56

Table of Contents

2.5.3	Dimension der Spezifität	56
2.6	Evidenzbasierte Feedbacksysteme zur Förderung von Schreibkompetenzen.....	58
2.6.1	Criterion.....	59
2.6.2	Writing Pal.....	61
2.6.3	CohViz.....	63
2.7	Forschungsdesiderate und weitere Entwicklung.....	68
3	DOES COMPUTER-BASED FEEDBACK FOSTER EXPOSITORY WRITING? ..	70
3.1	Introduction.....	72
3.2	Expository Writing	73
3.3	Feedback to Support Revision Processes.....	73
3.4	Computer-Based Feedback – A Feasible Approach to Foster Expository Writing? 74	
3.5	Potential Boundary Conditions of Computer-Based Feedback	76
3.5.1	Does the Specificity of Feedback Matter?.....	77
3.5.2	Does the Representation Format of Feedback Matter?	77
3.5.3	Does it Matter on Which Level of Text Quality Feedback is Provided?	79
3.5.4	Does Students’ Prior Knowledge Matter?.....	79
3.6	Overview of the Present Meta-Analysis	80
3.7	Method	81
3.7.1	Identifying Relevant Studies	81
3.7.2	Coding	83
3.7.3	Research Design and Outcome Measures	84
3.7.4	Moderator Variables	84
3.7.5	Computing Effect Sizes and Statistical Analyses.....	87
3.7.5.1.	Not Reported Data and Sensitivity Analysis	87
3.7.5.2.	Average Effect	88
3.7.5.3.	Publication Bias	88
3.7.5.4.	Used Data Management Tools.....	88
3.8	Results.....	89
3.8.1	Sample	89
3.8.2	Publication Bias.....	93
3.8.3	Overall Effect Size for Writing Quality	94
3.8.4	Heterogeneity Analyses.....	99
3.8.5	Moderator Analyses.....	99

3.9	Discussion.....	105
3.9.1	Summary of Evidence	105
3.9.2	What are the Theoretical and Practical Contributions of our Meta-Analysis? 107	
3.9.3	Limitations.....	108
3.9.4	Conclusion.....	109
4	THE MORE, THE BETTER?.....	111
4.1	Introduction.....	113
4.2	Strategy Instruction.....	114
4.3	Feedback	115
4.4	Combining Strategy Instruction and Formative Feedback	116
4.4.1	Underlying Processes when Combining Strategy Instruction and Feedback 118	
4.4.2	Sequence of Learning Phase and Practice Phase.....	119
4.5	Overview of the Present Study	121
4.6	Experiment 1	122
4.6.1	Research Questions and Hypotheses	122
4.6.2	Method.....	124
4.6.2.1.	Participants.....	124
4.6.2.2.	Design	124
4.6.2.3.	Materials	125
4.6.2.4.	Procedure	135
4.6.3	Results	136
4.6.4	Discussion.....	144
4.7	Experiment 2.....	145
4.7.1	Method.....	146
4.7.2	Results	148
4.7.3	Discussion.....	154
4.8	Experiment 3.....	155
4.8.1	Method.....	156
4.8.2	Results	158
4.8.3	Discussion.....	164
4.9	General Discussion	164
4.9.1	Summary of Evidence	165

Table of Contents

4.9.2	Theoretical Implications	166
4.9.3	Practical Implications	168
4.9.4	Limitations and Future Research.....	169
4.9.5	Conclusions	170
5	GENERAL DISCUSSION.....	171
5.1	Summary of Evidence.....	174
5.2	Theoretical Implications: Revised CAFCALI Model.....	177
5.2.1	Indications Regarding Orchestration.....	179
5.2.2	Indications Regarding Specificity	180
5.2.3	Indications Regarding the Level of Complexity	180
5.2.4	Indications Regarding Elaboration.....	181
5.2.5	Indications Regarding Prior Knowledge	182
5.2.6	Indications Regarding Arousal.....	182
5.2.7	Neglecting Factors: Sequence, Representation, and Cognitive and Metacognitive Processes	183
5.3	Future Research to Further Refine the CAFCALI Model	186
5.4	Practical Implications	189
5.4.1	Conscious Choice of a Feedback Tool.....	189
5.4.2	Conscious Choice of Complexity and Orchestration	190
5.5	Conclusion	192
	References	194
	List of Figures.....	236
	List of Tables.....	237
	Appendix	238
	Declaration on Contributions	249

List of Abbreviations

AI – Artificial intelligence

AUC – Answer until correct

AWE – Automated writing evaluation

CAFALI – Computer-assisted feedback-, context-, and learner-related impact on learning

EF – Elaborated feedback

I-PS – Instruction followed by problem solving

ITFL – Interactive two feedback loops

ITS – Intelligent tutoring system

KCR – Knowledge of correct response

KR – Knowledge of result

LP – Learning phase

MISCA – Message, implementation, student, context, agents

PP – Practice phase

PRISMA – Preferred Reporting Items for Systematic Reviews and Meta-Analyses

PS-I – Problem solving followed by instruction

RQ – Research question

1 GENERAL INTRODUCTION AND THEORETICAL BACKGROUND

When learners have problems facing the complex process of learning, they might use, for instance, only superficial learning strategies, resulting in no deeper understanding and missed opportunities to connect new information with already existing knowledge structures (see OECD, 2023). Therefore, support is needed to assist learners in their learning processes and in developing and applying learning strategies. Feedback is considered a beneficial instructional strategy for supporting learners during their learning processes (Hattie & Timperley, 2007; Shute, 2008; Wisniewski et al., 2020). However, it is laborious and time-consuming to provide individualized feedback, particularly in heterogeneous classrooms or individual learning settings (Lipnevich & Panadero, 2021; Shute, 2008). Computer-based feedback technologies have been explored as a useful addition to instructor feedback, as educational technologies are seen as a helpful supplement in provision of immediate and individualized feedback (Narciss, 2013). However, there is still a lack of evidence on the effectiveness of computer-based feedback (Mertens et al., 2022). Previous research has produced heterogeneous findings regarding both the implementation and orchestration (e.g., with prior instruction) of such feedback (Fleckenstein, Liebenow, & Meyer, 2023; Fyfe & Rittle-Johnson, 2016; Graham et al., 2015; Salden et al., 2010; Strobl et al., 2019). Furthermore, it is unclear whether specific conditions limit the effectiveness of computer-based feedback (Fleckenstein, Reble et al., 2023; Lipnevich & Panadero, 2021; Narciss, 2008; Strobl et al., 2019).

In this dissertation, the aim was to investigate the following overarching question: What are the factors that significantly influence the effect of computer-based feedback on learning?

Addressing this question, considering the context, and underlying processes, will help scholars and practitioners understand how computer-based feedback should be provided so

GENERAL INTRODUCTION AND THEORETICAL BACKGROUND

that it has a positive, supportive effect on learning. To achieve this aim and answer the research question, this dissertation contains three articles (two already published in international journals; one under major revision). The first is a theoretical paper that presents an overview of the literature and the existing computer-based feedback writing tools as well as elaborates upon factors that might affect the influence of computer-based feedback on learning. The second article is a pre-registered meta-analysis that examines the effect of computer-based feedback on fostering students' expository writing and empirically investigates the theoretically driven influencing factors from Article 1. The third article presents three pre-registered empirical experiments on the use of computer-based feedback to foster students' problem solving in the domain of physics and investigates the combination of strategy instruction and computer-based feedback.

The innovative added value of this dissertation is that, in contrast to earlier studies where different types of feedback were often investigated together (peer feedback, instructor feedback, computer-based feedback), this dissertation focused only on formative and computer-based feedback in order to investigate the effectiveness of the latter. To accomplish this, I used distinct perspectives and methods. For instance, this dissertation included different subject domains (writing and physics) and two ways of generating computer-based feedback (system-generated and instructor-generated/computer-mediated). Furthermore, the factors that can affect the effect of computer-based feedback were investigated in different ways. The key value of this dissertation is its holistic focus, which considers feedback-related, learner-related, and context-related factors and embeds them in one overarching model, named the CAFCALI (Computer-Assisted Feedback-, Context-, and Learner-related Impact on Learning) model. Beyond looking at feedback-related factors (e.g., representation, elaboration, specificity, complexity of content) in different subject domains, I investigated a wide range of underlying mechanisms by including (previously under-researched) learner-related factors (i.e., prior knowledge, and affective-motivational,

cognitive, and metacognitive processes). In addition, I investigated context-related factors (i.e., orchestration with prior instruction, sequence). This describes a second key value of this dissertation, as I contribute to the limited but necessary research on the combination of computer-based feedback and instruction and provide first insights into the orchestration of different types of computer-based feedback by combining them with strategy instruction.

The following chapters describe the theoretical backgrounds of the constructs needed to understand and investigate the research question of this dissertation. The independent variable in all of the studies conducted in the framework of this dissertation is learning outcome. This is why Chapter 1.1 briefly outlines the concept of knowledge (gain) and learning in general and explains why students may need support with that process. Chapter 1.2 contains a definition of feedback and a theoretical conceptualization of the feedback process. Based on this, Chapter 1.3 presents the specialty of computer-based feedback. Chapter 1.4 focuses on the CAFCALI model. Chapter 1.5 outlines, based on previous research (gaps), the goal and overarching research question of this dissertation. Chapter 1.6 provides an overview of the studies conducted as part of this dissertation, which follow in Chapters 2 to 4. Last but not least, Chapter 5 provides a summary and discussion of the study results and answers the overarching research question (Chapter 5.1), lists the theoretical implications based on the revised CAFCALI model (Chapter 5.2), and provides suggestions for future research (Chapter 5.3), as well as implications for practice (Chapter 5.4). Finally, the dissertation ends with a conclusion in Chapter 5.5.

1.1 Learning

Learning is a multifaceted and very broad term that is defined differently in different disciplines. In doing so, each discipline has a different area of reference in mind (e.g., psychological, physiological, sociological, educational). In general, learning can be modeled as a specific type of change in cognitive systems, including in those of human individuals.

Learning

Accordingly, learning can explain why a certain change has occurred (Schmidt, 2003). For instance, one might notice that someone's performance has changed (e.g., from pretest to posttest) and attributes this to the fact that the person now knows or can do more than before, which means the person must have learned something. This example shows that learning describes a process, i.e., the phase from before learning (the starting point) to after learning (the learning outcome or new starting point for further learning). This process can be viewed from different perspectives (Schmidt, 2003). Regarding a dimension of time, a learning process can be short-term (e.g., one teaching lesson, an immediate posttest) or long-term (e.g., one school year, a delayed posttest). Regarding a social dimension, learning can take place individually or organizationally. Regarding a factual dimension, learning can take place at the process level, can change the level of awareness, or can relate to interaction or to self- or external reference. Distinctions must also be drawn between the learning process itself, the action perspective (learning goals), and the observation of learning (learning results), i.e., who decides whether a learning process has taken place or not, oneself (self-reference) or an external observer (external reference). How to measure whether a learning process has taken place depends upon which disciplinary perspective is taken or which paradigm of educational theory is used and how, for example, knowledge is modeled therein.

1.1.1 Types of Knowledge

From a cognitive psychological perspective, conceptual knowledge (i.e., “knowing that” and “knowing how”) can be distinguished from skills (i.e., what a person can actually do). Regarding conceptual knowledge, de Jong and Ferguson-Hessler (1996) drew a distinction between declarative knowledge (“I know that ...”; also called lower-order or factual knowledge) and procedural knowledge (“I know how”; also called higher-order or inference knowledge; see also Anderson & Schunn, 2000; Baartman & de Bruijn, 2011; Jacob, 2022; Renkl, 2009; Wharton & Kintsch, 1991). Declarative knowledge describes superficial

factual knowledge that a learner could just memorize and reproduce without necessarily understanding the meaning (Anderson & Schunn, 2000; Baartman & de Bruijn, 2011; Gräsel et al., 1997; Gruber et al., 2000). Procedural knowledge requires that a learner interpret the information correctly and connect it with prior knowledge (pieces of declarative knowledge), which allows the learner to transfer the information to another situation or related content (Baartman & de Bruijn, 2011; de Jong & Ferguson-Hessler, 1996). The amount of transfer can be described using the terms near transfer and far transfer knowledge. For example, with near transfer knowledge, learners are able to transfer the learned information to a very similar situation, such as a task with different values but in the same domain. Far transfer knowledge enables learners to use the learned information to solve a problem in, for example, another domain. It requires thinking further and interpreting the information correctly (Jacob, 2022; Renkl, 2009; Weinstein & Mayer, 1986).

Declarative and procedural knowledge can be further differentiated into domain-specific, cross-disciplinary, and metacognitive knowledge (van Berkum & de Jong, 1991). An example of domain-specific declarative knowledge, knowledge in a specific subject or on a subject matter, is knowing the formula needed to calculate the current in an electric circuit, while domain-specific procedural knowledge would involve applying the strategy and calculating the current (e.g., using the formula). Cross-disciplinary declarative knowledge would include, for instance, knowledge about general argumentative structures; cross-disciplinary procedural knowledge would be to argue. Metacognitive knowledge is also a rather cross-disciplinary construct comprising knowledge about the task, the context, and the knowledge (acquisition) itself on a meta-level (Krathwohl, 2002). Examples of declarative metacognitive knowledge include knowledge about the utility and use of learning strategies or knowledge about a specific type of task. Procedural metacognitive knowledge means, for instance, planning how to apply a learning strategy, the regulation of cognitive processes, or the monitoring of one's own understanding (Renkl, 2009).

Learning

Within the context of learning, the term competence is also widely used. Many researchers define competences as holistic conceptions comprising integrated pieces of different facets of knowledge, skills, and attitudes, that enable persons to solve tasks (Baartman & de Bruijn, 2011; Renkl, 2009). For example, writing competence or writing literacy describes the ability to understand the function of writing in the lifeworld, to communicate clearly (via text), and to use writing as a tool in one's daily life and work. Baartman and de Bruijn (2011), however, postulated that the competence itself, in the sense of integrated knowledge and skills, is the actual learning product and that the integration of the various pieces of knowledge and application of those pieces of knowledge and skills constitutes the learning process, because learning is the processing of information. In other words, one does not simply receive information and then "know" it. Instead, information meaningful only when it is interpreted on the basis of prior knowledge (Renkl, 2008).

1.1.2 Acquisition, Construction, and Integration of Knowledge

Schmidt (2003) distinguishes between the acquisition of knowledge, reflection on one's own knowledge (applying metacognitive knowledge) and the application of knowledge (performance), and the observation of these steps from an external perspective (measuring knowledge acquisition and application and identifying it as learning). From a cognitive psychological perspective, this highly complex operation of knowledge acquisition and processing (Renkl, 2009) takes place in different memories (Atkinson & Shiffrin, 1968). First, in the sensory memory, new information is collected; second, some of this information is coordinated and processed in the working memory (Baddeley, 1992). As the capacity of the working memory is limited, any information that is considered less important gets forgotten, and the information that is considered important is transferred in a third step to the long-term memory. In the long-term memory, new information is connected with prior

knowledge and is integrated in a structured knowledge scheme (Jacob, 2022; Mayer, 1996; Wittrock, 1974).

Knowledge schemes represent connections between problem solving procedural knowledge, domain-specific information, and knowledge about a concrete situation (Kintsch, 1988). Hereby, Wharton, and Kintsch (1991) distinguish between knowledge construction and knowledge integration. Knowledge construction describes the process when through which single words of newly provided information are classified in linguistic sense units (cf. linguistic representation), which in turn are connected to each other within a sentence, creating a network of statements (cf. propositional network). This can then be interpreted and elaborated upon with prior knowledge (cf. elaborated propositional network). Next, the knowledge should be integrated to refine this elaborated network by structuring and classing it with other content or demarcating it from different content. This means that when a person is involved in a similar problem situation in the future and wants to transfer the knowledge, the existing knowledge schemes can be activated and populated with the updated concrete details. In this way, learners can develop problem solving strategies that are applicable in related contexts (Renkl, 2009; VanLehn, 2008). Learning strategies can help learners with knowledge construction and integration as they specifically trigger these cognitive processes. For example, generative learning helps learners comprehend an explanation (e.g., an instruction at the beginning of a teaching unit) and interpret it, as learners must generate active connections between the contents to be learned and their prior knowledge, which leads to understanding (Renkl, 2009; Wittrock, 1990).

Besides the cognitive processes that take place in the memory and the applications of metacognitive knowledge to regulate these cognitive processes and monitor understanding of the learning content, there are also motivational conditions that influence learning. Learning motivation describes the wish or intention to acquire specific content or competences (Möller, 2008). Motivation can be categorized as intrinsic (when one is

Learning

interested in the topic or because the learning activity makes fun) or extrinsic (learning to avoid sanctions or to receive praise or rewards). Furthermore, different aspects of the learners' subjective perceived self-determination of learning can affect students' motivation (Ryan & Deci, 2000). Research shows that self-determined and intrinsic aspects of motivation foster deeper understanding (Ryan & Deci, 2000).

Here, the pedagogical understanding of learning comes into play. According to that understanding, specifically that learning always means an interaction with the social world in which the learner thinks and acts (Billett, 2001; Göhlich et al., 2007), the interpretation of new information is influenced by, for example, the environment, life experience, and education of the learner. Pedagogically, knowledge extends beyond mere internal memorization and categorizing into declarative and procedural types of knowledge. Instead, it encompasses a broader spectrum, viewing learning as a fundamental step towards engaging and acting within a social context (Baartman & de Bruijn, 2011; Billett, 2001).

To sum up, first, the term learning is defined differently in different disciplines, and second, it is affected by various factors. From a pedagogical point of view, learning is inseparably connected to the world the learner lives in and is omnipresent, since a person cannot not learn, but can do so implicitly without being aware of it. Learning therefore means integrating the knowledge and skills of an individual in the world, thereby developing the world and oneself further. Learning thus enables people to act. From a psychological point of view, the understanding of learning focuses on the internal processes that take place when learning. Here, the cognitive load, for example, plays a big role.

In general, it can be said that learning is a complex, active, and individual process that can differ depending on the subject domain, requires learning strategies, and is affected by cognitive, metacognitive, and affective-motivational factors (Renkl, 2009). In this dissertation, learning is defined as the activity of processing information with the aim of being able to solve problems yourself. The learner has an active role (e.g., deciding whether

to read a text, which parts of it to read, how quickly, what answers to give, or which solution strategy to choose, etc.), but the instructor provides the content. In order to measure learning progress, this dissertation focuses on short-term, individual learning in the form of a change in performance at the process level (from pretest to an immediate posttest with integrated phases of knowledge acquisition and knowledge application) as observed from an external perspective and in the context of formal learning. Knowledge acquisition and processing are highly cognitive-demanding processes that require different types of knowledge and depend on cognitive, metacognitive, and affective-motivational factors. This may lead to learners facing difficulties in learning, as a result of which only superficial understanding occurs and new information is not linked to already existing knowledge structures, particularly on an individual process level (cf. Sweller, 1994). Thus, learners need support at the process level; in this dissertation, that support is provided through feedback.

1.2 Feedback

When investigating feedback, researchers have taken a variety of different methodological perspectives, used different terminology, and based their research on different theoretical models (cf. jingle-jangle fallacy; Block, 1995). Originally, feedback was an industrial term that described a loop of information that was generated by an output and communicated back to the system (Lipnevich & Panadero, 2021). Today, feedback is an everyday term used in many different contexts, and it can play a key role in learning.

In her review, Shute (2008) emphasized that feedback, in computer-based environments as well as in classrooms, follows the input of the learner “with the purpose of shaping the perception, cognition, or action of the learner”, (Shute, 2008, p. 175) and with the overall aim of enhancing “learning, performance, or both, narrowing the formation of accurate, targeted conceptualizations and skills” (Shute, 2008, p. 175). Feedback can be used, for instance, for evaluation purposes or assessments, such as assigning grades, as well as for

Feedback

diagnostic purposes, such as controlling whether a learning objective is achieved or reactivating students' prior knowledge by providing information about a learning strategy to solve a task (Azevedo & Bernard, 1995; Narciss & Huth, 2004; Shute, 2008).

The feedback information is based on the learning goals, standards, or various success criteria (Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Narciss & Huth, 2004). The success criteria can be explicitly (e.g., "I want you to write an essay with high cohesion on the topic ...") or implicitly (e.g., "I want you to write a good essay on the topic ...") addressed in the task. However, feedback includes not only assessment (or evaluation) of the performance against the success criteria but also cues on how to reduce the discrepancies (e.g., providing solution steps; Stevenson & Phakiti, 2014; Winne & Butler, 1994). Depending on the goal, different types and forms of feedback may be used; these may convey specific information and messages to the learner (Hattie & Timperley, 2007; Mertens et al., 2022; Winne & Butler, 1994), and they can aim at different levels (e.g., task level, process, level; Hattie & Timperley, 2007).

Regarding the types of feedback, Hattie and Timperley (2007) proposed three types of feedback in their feedback model: feed up, feed back, and feed forward. Feed up is based on the learning goal and related to the success criteria by which completion of the task can be measured. Feed back (generally understood as feedback) means providing information about the answer to a question or about the current state of performance in relation to the goal that is being targeted (e.g., whether the task was solved correctly or incorrectly). Feed forward addresses the question, "Where to next?" by providing further information about the solution to the task (e.g., by presenting solution steps or suggestions), thus providing the learner with greater learning opportunities.

Furthermore, following Hattie and Timperley (2007), feedback can be given on four levels: the task level (i.e., the level of current performance), the process level (i.e., the level

of the incorporated learning processes), the self-regulation level (i.e., the metacognitive level of regulation and control), and the self level (i.e., the level of personal evaluation).

The information communicated in the feedback is dependent on when the feedback is provided: One can distinguish between formative and summative feedback (Allen et al., 2016; Hattie & Timperley, 2007; Scriven, 1967). Formative feedback occurs during the learning process, for example, feedback on a first draft, which is provided so that the learner can process and implement it during revision phase, as well as possibly get a second round of feedback on the revision. The primary goal of formative feedback is to guide the learner in regulating and enhancing subsequent phases of the learning process (Butler & Winne, 1995; Hattie & Timperley, 2007; Narciss, 2008, 2012; Shute, 2008). Summative feedback occurs at the end of task performance, for example, grades at the end of a school year (Harrison et al., 2015; Shute, 2008).

In addition to the type, the level, and the timing of feedback, there are various other dimensions by which feedback can be conceptualized (Lam et al., 2011), as use (uptake, implementation) by the recipient depends on multiple properties of the feedback, such as specificity (e.g., Goodman et al., 2004) and source (e.g., Greller & Herold, 1975; Vancouver & Morrison, 1995), as well as on personal traits of the recipient, including learners' arousal (e.g., Kluger & DeNisi, 1996; Kluger et al., 1994) and students' prior knowledge (Narciss & Huth, 2004; Patchan et al., 2016). Several feedback models purpose unique conceptualizations of the feedback process.

1.2.1 Feedback Models

On the one hand, there are rather behavioristic models (cf. Bangert-Drowns et al., 1991; Kluger & DeNisi, 1996; Kulhavy & Stock, 1989) in which corrective feedback to a response is assumed to result in a change in behavior. On the other hand, there are models that argue for feedback as a learning opportunity, not only to change students' behavior but also to

Feedback

support them in constructing knowledge, and in these models, students have an active role in processing the feedback (Sadler, 1989; Shute, 2008). Researchers saw the importance of feedback engagement and developed models in which the learner is not only seen as a feedback receiver, but also as a feedback provider, for instance, feedback to oneself or in peer feedback situations (Kluger & DeNisi, 1996; Narciss & Huth, 2004; Nicol & MacFarlane-Dick, 2006; Sadler, 1989). In the field of educational psychology, more cognitive models are examining the underlying mechanisms at play when one processes feedback and investigating their impact on students' performance (Butler & Winne, 1995; Kluger & DeNisi, 1996). Formative assessment, based on Black and Wiliam (1998), takes the view that, regardless of the instructional environment, feedback is informative for students as well as for the teacher (or another instructional agent). Based on this idea, scholars developed models in the context of self-regulated learning (Nicol & MacFarlane-Dick, 2006; Panadero et al., 2018) or combined with the control-value theory of emotions (Pekrun, 2017). As a further distinction, there are more general pedagogical approaches (Hattie & Timperley, 2007; Nicol & MacFarlane-Dick, 2006) but also models that focus on internal processes (cognitive, metacognitive, affective-motivational; Kluger et al., 1994; Narciss, 2008; Narciss & Huth, 2004) or on giving feedback in specific contexts, for instance, in computer-based settings (Narciss, 2008, 2013).

Based on this large variety of conceptualizations, feedback models can be said to vary in terms of the feedback definition, typology, and thematic foci and to focus on many different aims and aspects of the feedback process. Which feedback model is suitable depends on the goal of the research.

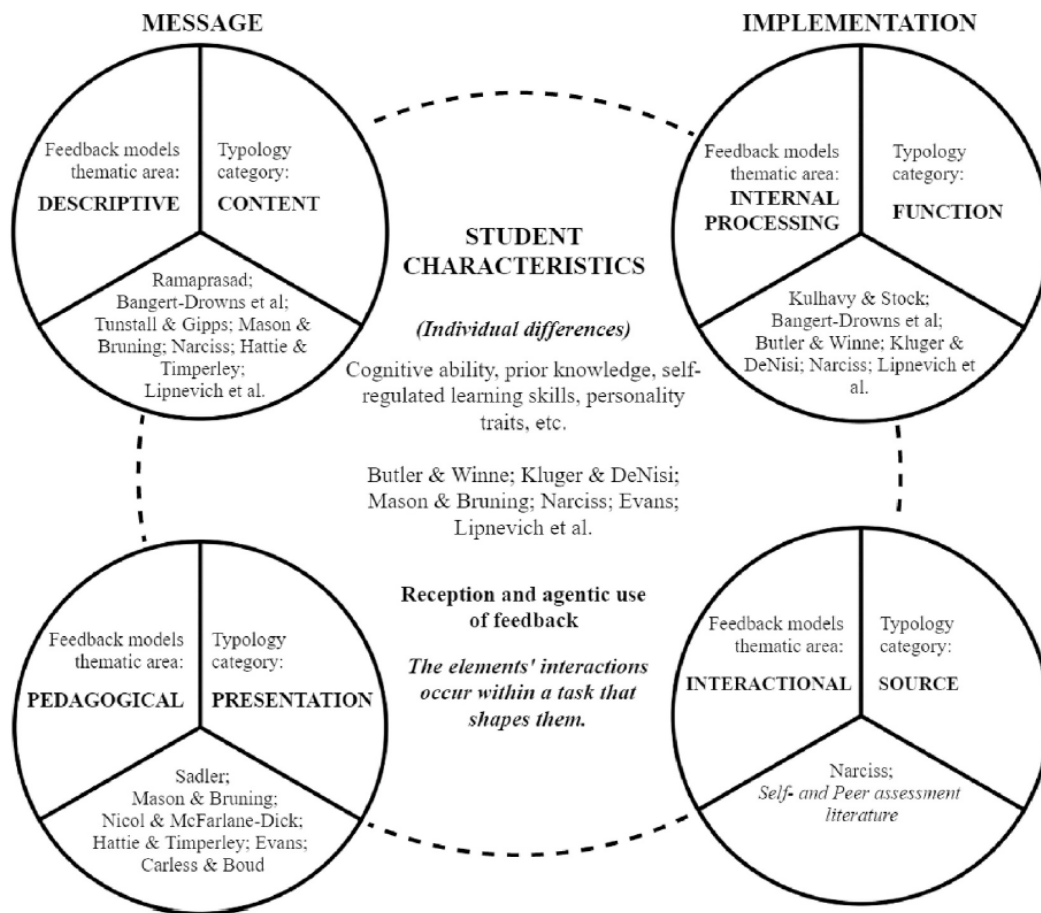
The wide range of models shows that various discourses were being held in parallel. However, there are also recognizable overlaps. In their review, Lipnevich and Panadero (2021) analyzed 14 feedback models and identified common key elements of the definitions that the models were based on. *Information* was a key component in all of the feedback

models as without information there is nothing to implement or process. Hereby, some argued that feedback is simply the concrete information that is provided to the learner (see, e.g., Kulhavy & Stock, 1989); others that it is only feedback when the information is processed by a learner and view the uptake of feedback as an interactive process (see, e.g., Bangert-Drowns et al., 1991). A second crucial component in most of the models was the *gap* between the current and the aspirational performance, as feedback intended to close this gap. Third, the *processing* of the feedback including cognitive, metacognitive, affective-motivational, or self-regulated processes, was mentioned in several models. Another element in more recent models was the *agent* providing the feedback, which ranged from peers to teachers to computer-based systems. Finally, the researchers agreed that, at least in the field of formative assessment, feedback is part of a complex process that includes more than the provision of static information to learners (e.g., “your answer is correct”); it also requires the learners’ active engagement. Therefore, they designated *students’ characteristics* as another element in their feedback definition.

In 2022, Panadero and Lipnevich developed a new model called MISCA (massage, implementation, student, context, agents; see Figure 1) that integrates the core elements of the 14 models in their review (Lipnevich & Panadero, 2021). The MISCA model is based on the classification of feedback models proposed by Narciss and Huth (2004): *content*, *function*, and *presentation* of feedback. Panadero and Lipnevich (2022) expanded this classification by the categories “reception” (students’ characteristics) and “source” of feedback, then classified the previous models within the MISCA model.

Figure 1

MISCA model by Panadero and Lipnevich (2022)



Message describes the content of the feedback it is the actual information that is provided to the recipient. The feedback message can have various formats according to the aim of the feedback (e.g., error-related, subject-related, guidance to support self-regulated strategies, etc.). As a general guideline, Panadero and Lipnevich (2022) proposed including the three types of feedback suggested by Hattie and Timperley (2007): feed up, feed back, and feed Forward. Providing learners with information on the goal, the current state, and how to close the gap between the current and the targeted performance, holds the potential to significantly enhance learning and performance. *Implementation* refers to students' internal processing of the feedback, that is decoding, analyzing, and implementing information according to the goal of the feedback, as well as students' subsequent internal cognitive, metacognitive, and

affective-motivational processes, and self-regulatory steps (Butler & Winne, 1995; Narciss, 2008; Sadler, 1989). Different types of feedback can contribute to varying functions of feedback, such as generating more correct responses, improving incorrect responses, self-regulation, and closing existing knowledge gaps (Mertens et al., 2022; Narciss, 2012), as well as providing feedback on task level, process level, self-regulatory level or self level (Hattie & Timperley, 2007). Panadero and Lipnevich (2022) placed *students' characteristics* as the central element of the model (cf. Brooks et al., 2021), considering the importance of the characteristics, differences, and specific needs of individual students (e.g., prior knowledge, self-efficacy, motivational beliefs, gender; Lipnevich et al., 2016). The *context* component includes the selection of a suitable pedagogical approach for delivering the feedback in terms of modality, instructional design, adaptivity, timing, amount of tries, appropriate presentation, domain-specific aspects (see also Narciss & Huth, 2004), as well as orchestration when combining feedback with other instructional strategies (e.g. prior instruction; Fyfe & Rittle-Johnson, 2016). Thus, the context of feedback is a crucial element the best formulated feedback message will not improve learning if the context is not considered (Hattie & Clarke, 2019). Last but not least, the *agent* emphasizes the source that provides the feedback (i.e., student themselves, teacher, peers, computer-based system). Finally, Panadero and Lipnevich (2022) highlighted the interactions between components of the MISCA model. For example, the feedback message can take different formats, depending on the agent providing the information, and the instructional context shapes the availability when selecting the agent.

1.2.2 Definition of Feedback

Regarding the definition and conceptualization of feedback, there seems to be common ground in the educational field. In this dissertation, feedback is defined based on the core elements identified by Lipnevich and Panadero (2021) in various theoretical feedback

models. Additionally, the key elements of the feedback process and its potential influencing factors, as summarized in the MISCA model (Panadero & Lipnevich, 2022), the ITFL-model (Narciss, 2006, 2008), the framework for designing formative feedback of Narciss and Huth (2004), and the fundamental review of formative feedback by Shute (2008), contribute to the comprehensive definition. In this dissertation feedback in general is defined as follows.

Feedback is post-response information that is provided by an external agent (e.g., peer, teacher, system, AI) or by internal sources (e.g., information perceived by the student when processing a task) and focuses on aspects of a learners' current performance or learning to help the learner to close the gap between the current performance or learning and the target performance or learning based on success criteria.

1.2.3 Empirical Evidence of Feedback

Feedback is considered an effective instructional strategy to support students in their learning processes and teachers in their teaching (Azevedo & Bernard, 1995; Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Mertens et al., 2022; Narciss, 2012; Shute, 2008; Wisniewski et al., 2020). To date, there has been a plethora of research on feedback in different instructional contexts and meta-analytical evidence that has contributed to the conventional wisdom that feedback is one of the most influential factors in diverse instructional settings, including digital learning environments (e.g., Bangert-Drowns et al., 1991; Hattie & Timperley, 2007; Mertens et al., 2022; van der Kleij et al., 2015).

For instance, Graham et al. (2011) conducted a meta-analysis ($k = 16$ studies) that focused on the overall effect of different forms of formative feedback (e.g., by peers, instructor; verbally, written, computer-based) on the writing quality of students from grades 2 to 9. The findings showed a moderate to large significant positive effect ($d = 0.77$, CI 95% [0.49, 1.08]) of feedback on writing quality in comparison to an alternative treatment or a control condition without any feedback. However, the authors themselves stated that the results must

be interpreted with caution due to the great heterogeneity and that the number of studies does not allow a clear conclusion to be drawn regarding the effectiveness of different forms of feedback.

More recently, Wisniewski et al. (2020) conducted a meta-analysis over 435 empirical studies ($k = 994$ effect sizes, $N > 61,000$ students) that investigated the effects of feedback on students' learning. Based on random-effect models, the results revealed a medium overall effect ($d = 0.48$) of feedback on learning. The results also showed a large heterogeneity ($I^2 = 83.43\%$) between the studies. Moderation analyses clarified that this could partly be explained by the type of feedback information, as high-information (elaborated) feedback had a greater impact on learning ($d = 0.99$, $k = 42$ effects) than corrective feedback ($d = 0.46$, $k = 238$ effects) or reinforcement/punishment ($d = 0.24$, $k = 39$ effects). Another influencing factor was the direction of feedback who provided feedback to whom, as peer feedback (student to student) was most effective ($d = 0.85$, $k = 16$ effects), followed by teacher to student feedback ($d = 0.47$, $k = 812$ effects) and student to teacher feedback ($d = 0.35$, $k = 27$). When interpreting these effects, however, the number of included studies should be taken into account and put into perspective (fewer studies often lead to larger effects than do analyses with larger sample sizes [small-study-effect]; Sterne et al., 2000).

Kluger and DeNisi (1996) produced similar outcomes in their meta-analysis based on 607 effect sizes (extracted from 131 papers, $N = 12,652$ participants) that examined the variability of the effects of feedback interventions on performance, as the authors obtained an average-small to medium effect of $d = 0.41$. However, they also observed a large heterogeneity: Over a third (32%) of the effects had a negative influence on performance. Kluger and DeNisi (1996) concluded that feedback interventions are not always effective but depend on various boundary conditions; feedback interventions can reduce or improve performance. To identify the relevant boundary conditions for implementing feedback effectively, the authors performed several moderator analyses. The major findings showed

that praise had an attenuated effect on students' meta-task processes which led in many cases to reduced performance (see also Balcazar et al., 1985). In contrast, feedback interventions that supported students' attention to the task (regarding motivation or learning processes) produced higher and positive effects (this also applies for computer-based interventions) on students' performance (Kluger & DeNisi, 1996). Another main result of Kluger and DeNisi's (1996) meta-analysis was that the effectiveness of feedback interventions was found to be moderated by, for example, 1) the type of task as memory tasks produced higher effects than physical tasks or following rules, 2) the complexity of the tasks since complex tasks revealed weaker effects than less complex tasks, 3) the goal setting as feedback which included goal setting was more effective than that without, and 4) the elaboration level of the feedback information, as providing the correct solution was more effective than not providing it.

1.3 Computer-Based Feedback

In the case of computer-based feedback provided by a system, the software takes the role of the feedback providing agent, and it usually utilizes machine learning algorithms. Educational technologies are used to provide system-generated feedback via, for example, an algorithm or artificial intelligence (AI) that analyzes the student's draft of a text and then automatically reports back errors or offers suggestions for improvement (see, e.g., Dai et al., 2023; for examples of AI-based feedback using ChatGPT; see Montenegro-Rueda et al., 2023; Scheurer et al., 2022; Tsai et al., 2020; Wu et al., 2022). Computer-based feedback can also be person-mediated, which means that people, such as teachers or peers, provide feedback via an electronic platform. For example, a teacher can formulate a suitable response to various possible answer options for a multiple-choice question and the system then provides the corresponding feedback once an answer is chosen by a student.

Using computer-based settings and feedback tools in learning environments entails various advantages and disadvantages. Computer-based feedback often takes place in

individual learning phases, which means that teachers cannot control how students perceive and implement the feedback or how they react (e.g., emotionally, or motivationally) to the feedback (Alharbi, 2017). For example, students may feel less recognized, less supported on a social level, or less engaged when a system, instead of a human, provides correction and encouragement (Fong et al., 2019; Fong & Schallert, 2023). In addition, applying computer-based feedback requires digital literacy of teachers and students, as well as the necessary technical equipment. Furthermore, teachers' behavioral intention and attitude towards educational technologies are determined by teachers' perceived usefulness and the ease of using of digital technology in education (Scherer et al., 2015). However, when teachers are open to use computer-based tools and the (technical) requirements are met, applying computer-based feedback tools combines several advantages.

A great advantage of implementing computer-based feedback is that it can save time. Although feedback has been demonstrated to be effective in fostering learning, especially providing individual, formative feedback, is often time-consuming and laborious (Cho & MacArthur, 2010). Particularly in full classes it is difficult to pay the same amount of attention to everyone. Since teachers usually have little time to offer individual feedback to learners (Cho & MacArthur, 2010) and cannot provide immediate feedback to every student in a full class simultaneously, computer-based methods of automatic feedback are a particularly useful supplement to instructor feedback to support learning (Kellogg & Whiteford, 2009; Roscoe & McNamara, 2013; Strobl et al., 2019) and can support learners individually and immediately (Alruwais et al., 2018; van der Kleij et al., 2012, 2015). Even if teachers' capacity is limited, through computer-based feedback, all students in a class can receive immediate, individualized, and self-regulated feedback (Shute, 2008; Wisniewski et al., 2020). Teachers can also face problems when students spontaneously ask for feedback; computer-based technologies enable teachers to prepare feedback in advance. Another advantage is that computer-based feedback is more objective than human feedback, as the

same output is provided for the same input, and it is less biased by potential prejudices toward the students (Kluger & DeNisi, 1996; Shute, 2008). From a technological perspective, empirical findings have shown that computer-based systems are able to provide more accurate feedback than human raters on operational characteristics (see, e.g., orthography, local cohesion; e.g., Burkhart et al., 2020). Furthermore, computer-based tools increase the range of feedback strategies that can be implemented (Narciss, 2013) and allow for multimodal and multimedia feedback, which means, for example, if oral teacher feedback is not perceived or not perceived as intended, computer-based feedback can display and store the information in written or visual formats (Ainsworth, 2006; Sweller et al., 2011). It is also possible to integrate adaptive elements and provide different levels of support for students.

1.3.1 Empirical Evidence Regarding the Implementation of Computer-Based Feedback

In spite of the theoretical potential of computer-based feedback, meta-analytical studies on the use of computer-based feedback reveals a diverse picture regarding its general effectiveness. For example, Graham et al. (2011) conducted a meta-analysis of 16 comparisons from true-experimental and quasi-experimental studies to investigate the effect of feedback on overall student writing quality (from second to ninth grade; $N = 1011$ students). In the studies, students in the experimental condition received formative writing assessment, and those in control condition received another or no treatment. The authors conducted a sub-analysis that focused only on computer-based feedback based on four studies ($ES_1 = -.07$, $N = 835$; $ES_2 = .35$, $N = 111$; $ES_3 = .38$, $N = 243$; $ES_4 = .60$, $N = 52$). They found a small, non-significant overall effect of computer-based feedback on writing quality compared to business-as-usual interventions. However, these results were based on the synthesis of only four very different empirical studies with heterogeneous results, which limits the statistical significance of the results, particularly with regard to the sample sizes of the included studies and the significance of the effect. The problematic validity becomes

even clearer when comparing the results of Graham et al. (2015), who also considered four studies. Three were the same as in Graham et al.'s (2011) meta-analysis. By excluding the study by Shermis et al. (2004), which had the highest number of student participants ($N = 835$) who reported a negative effect ($ES = -.07$), and instead including a different study (Holman, 2011) that revealed a positive effect ($ES = .44$), they concluded that computer-based feedback is effective in improving students' writing quality with a significant medium positive effect of $d = 0.38$.

Mertens et al. (2022) conducted a network meta-analysis based on 163 effect sizes of 77 studies published between 1968 and 2021. The aim was to investigate the effectiveness of different computer-based feedback types (e.g., knowledge of results, knowledge of correct response, answer until correct, elaborated feedback) on learning outcomes primarily those of school students in different domains. Thereby, the authors differentiated between lower-order learning (e.g., recall; based on 115 effect sizes) versus higher-order learning (e.g., transfer; based on 48 effect sizes). In addition, due to the network meta-analysis approach the authors were able to rank the feedback types according to their effectiveness. Mertens et al. (2022) found small (Hedge's $g = 0.24$) to medium-to-large (Hedge's $g = 0.71$) positive significant effects for all types of computer-based feedback on lower-order learning outcomes and significant medium effects of answer until correct feedback and elaborated feedback on higher-order learning outcomes. Knowledge of results and knowledge of correct response feedback did not produce a significant effect on higher-order learning outcomes (Mertens et al., 2022). Thus, the authors provided evidence that the type of feedback (e.g., corrective vs. elaborated) and the type of learning outcome (lower- vs. higher-order) determine the effectiveness of the feedback. Their results revealed that elaborated feedback produced the largest effects on both lower-order ($g = 0.71, p < .01$) and on higher-order learning ($g = 0.46, p < .01$). In contrast, knowledge of correct response feedback was least

effective in supporting learning, and this feedback only affected lower-order learning ($g = 0.24, p = .026$), not higher order learning ($g = 0.14, p = .345$).

Another factor contributing to the limited understanding of the effectiveness of computer-based feedback is the lack of in-depth research on this topic. Previous research has predominantly focused on feedback in general including computer-based feedback as a subgroup with a limited number of heterogeneous studies (cf. Graham et al., 2011, 2015). Moreover, researchers have only rarely investigated the orchestration of computer-based feedback (e.g., its combination with prior instruction, which is often the case in practice).

1.3.2 Empirical Evidence Regarding the Orchestration of Computer-Based Feedback

Most instructional situations start by inducing knowledge through input or instruction. The aim of starting with the provision of a strategy instruction (for example, a procedural explanation of parallel electrical circuits, Ohm's law, and how to apply it) is that learners build up a consistent mental representation for that type of problem and the corresponding solution strategies (e.g., how to fill out the formula to calculate the electric current in an electrical circuit). In subsequent practice phases, students apply and practice the acquired knowledge during a subsequent practice phase, in which they can receive feedback on their performance. Here, the provided feedback is designed to further support students in applying the previously learned strategy to solve the task and in transferring the knowledge to solve different tasks (higher-order learning). Considering that both strategy instruction and feedback have been shown to be effective in supporting student learning, they are frequently combined in educational settings (Magliaro et al., 2005; Rosenshine, 2008). From an alternative perspective, providing strategy instruction in advance carries the risk that students may underestimate the specific task requirements during the practice phase, resulting in reduced processing of the feedback (Narciss, 2008; see also illusion of understanding, Renkl, 2014).

Regarding computer-based settings, initial research found that the combination of computer-based feedback and computer-based instruction reduced learning (or had no additive effect) compared to computer-based instruction programs without computer-based feedback (Carroll & Kay, 1988; Fyfe & Rittle-Johnson, 2016; Kluger & DeNisi, 1996; Lepper & Gurtner, 1989; Wise et al., 1989).

Salden et al. (2010) compared eight studies across different domains that combine strategy instruction (i.e., worked examples delivered by a cognitive tutor) and problem solving situations that include different types of feedback (e.g., context-sensitive hints, immediate feedback, and prompts). The findings showed that the combination of strategy instruction and tutored problem solving enhanced learning. In contrast, Fyfe and Rittle-Johnson (2016) conducted experimental studies with elementary students to investigate the combined effects of computer-based prior strategy instruction in a learning phase and computer-based feedback in a subsequent practice phase on students' mathematical problem solving. The strategy instruction comprised an explanation of how to use the equalize strategy and a demonstration of four different math equivalence problems that could be solved by applying the correct problem solving strategy. The feedback was presented in the form of knowledge of result feedback and praise (e.g., "Good job! You got the right answer" or "Good try, but you did not get the right answer"; Fyfe & Rittle-Johnson, 2016, p. 85). In Study 1 ($N = 108$), they crossed the factors strategy instruction (yes vs. no) and immediate corrective feedback (yes vs. no). The researchers identified a strategy-by-feedback interaction, suggesting that students' learning was enhanced by to strategy instruction only when it was not combined with feedback. In Study 2, all students received prior strategy instruction and 12 practice tasks in a subsequent practice phase. The researchers manipulated which type of feedback (no vs. immediate vs. summative feedback) the students received after each practice task. As in Study 1, students benefited more from the strategy instruction in the no-feedback condition than in the combined condition in which they received both

prior strategy instruction and additional feedback. McLaren et al. (2008) found a similar pattern of the combination of immediate feedback and worked examples within a cognitive tutoring system for chemistry. Additionally, Wischgoll (2017) showed related evidence regarding a reduced interaction effect on students' writing quality in higher education when peer-given strategy instruction (training on writing strategies, i.e., applying text structures, language use, summarization) was combined with informative tutoring feedback or try-again feedback in a computer-based learning environment. More recently, Shirah and Sidney (2023) conducted an experimental study based on a 3-by-2 design with 171 undergraduate students. The authors aimed to investigate whether it made a difference if students were induced with relevant or irrelevant prior knowledge in combination with no feedback, immediate computer-based feedback, or delayed computer-based feedback. They found that immediate feedback supported learning more than no feedback only when irrelevant prior knowledge had been induced. Immediate feedback combined with topic-related prior knowledge activation had no effect on learning.

In total, these contradictory findings offer initial evidence on the orchestration of feedback in combination with advance induced strategy knowledge indicating that the combination is not necessarily effective at fostering students' learning and can also reduce the effect of single interventions (see also Kluger & DeNisi, 1996; Narciss, 2008). At the same time, previous research has indicated that the effectiveness of feedback strongly depends on learners' prior knowledge. It remains unclear what additional factors might influence computer-based feedback.

1.3.3 Potential Influencing Factors on the Effectiveness of Computer-Based Feedback

According to the theoretical models of the feedback process, various influencing factors must be considered during both implementation (providing feedback) and uptake (processing) to ensure that feedback is beneficial for learning. Shute (2008) criticized that

there are gaps in the feedback literature regarding feedback in instructional contexts, the potential mediating impact of student characteristics, and potential interactions of feedback and task characteristics (see also Strobl et al., 2019).

According to the MISCA model, potential influencing factors may include the content of the feedback (cf. lower- and higher-learning outcome), the implementation and function of the feedback (e.g., specificity, corrective vs. elaborated), students' characteristics (prior knowledge, cognitive, metacognitive, and motivational processes), the representation and context of feedback, and the source (considering the computer as a possible agent). In contrast to the rather general MISCA model, Narciss' (2006, 2008) Interactive Two Feedback Loops Model (ITFL-model) explicitly refers to computer-based feedback. Similar to the work of Narciss and Huth (2004), Narciss (2008) emphasized the importance of considering how computer-based feedback is embedded in the context and highlighted the role of individual cognitive, metacognitive, and motivational processes. The ITFL-model treats feedback implementation as a mutual interaction involving external (feedback) and internal (learner) factors that has a potential impact on the effectiveness of feedback (cf. internal and external feedback loop; see also Butler & Winne, 1995; Hattie & Timperley, 2007; Panadero & Lipnevich, 2022, for related assumptions).

Based on the framework by Narciss and Huth (2004), MISCA (Panadero & Lipnevich, 2022), the ITFL-model (Narciss, 2008) and on previous literature and research, the influencing factors on the process of computer-based feedback can be summarized as follows:

- Feedback-related factors include the complexity of the content of the task and the feedback (cf. lower- and higher-learning outcomes; Mertens et al., 2022; Panadero & Lipnevich, 2022; Strobl et al., 2019), specificity of feedback (Goodman et al., 2004; Moreno, 2004; Shute, 2008), level of elaboration (Mertens et al., 2022;

Strobl et al., 2019), and representation of the feedback (Ainsworth, 2006; Butler, 1987; Wiliam, 2007).

- Learner-related factors include prior knowledge (Fyfe & Rittle-Johnson, 2017; Mory, 2004; Narciss & Huth, 2004; Shirah & Sidney, 2023), cognitive processes (Butler & Winne, 1995; Sweller et al., 2011), metacognitive processes (Baars et al., 2013; Finn & Metcalfe, 2014; Narciss & Huth, 2004), and affective-motivational processes (Kluger et al., 1994; Kuklick & Lindner, 2023).
- Context-related factors include instructional orchestration of feedback (Fyfe & Rittle-Johnson, 2016; Salden et al., 2010; Wischgoll, 2017), and the timing of the feedback (i.e., sequence of the orchestration).

1.3.3.1. Do Feedback-Related Factors make the Difference?

Specificity of feedback concerns how clearly and in how much detail the feedback information is presented (Goodman et al., 2004; Lipnevich & Panadero, 2021; Patchan et al., 2016). One can distinguish between specific feedback, which provides the learner with, for example, an explanation of what was done wrong when solving the problem or directly highlights the error in the learner's response, and generic feedback, which might include, for example, holistic feedback on the performance in general. Previous studies have demonstrated that specific feedback contributes to improvements in learning (Hattie & Timperley, 2007; Lachner et al., 2017a; Patchan et al., 2016; Shute, 2008; Wisniewski et al., 2020). Most previous research has shown that generic feedback is perceived as more challenging to implement and can frustrate learners which may consequently lead to lower learning gains (Kluger & DeNisi, 1996; MacArthur, 2015; Moreno, 2004; Patchan et al., 2016; Roscoe & McNamara, 2013; Shute, 2008; Sturm, 2014, 2016).

Representation of feedback refers to its visual presentation. Besides spoken feedback, there are, for instance, numerical representations that are frequently presented in the form of

scores, points, percentages, grades, charts, rating scales, or diagrams (Kellogg et al., 2010; Roscoe, Snow, & McNamara, 2013; Wilson & Roscoe, 2020). Furthermore, feedback can be visualized in pictorial formats, such as in the form of concept maps or with emojis, symbols, or arrows (Kim, 2012; Lachner et al., 2017a, 2017b; O'Rourke et al., 2011; Pirnay-Dummer & Ifenthaler, 2011; Villalon & Calvo, 2011). In addition, there are also embedded formats, such as signaling, highlighting, or marking errors or important information (Burkhart et al., 2020; Moore & MacArthur, 2016; Tsai et al., 2020). In some cases, it makes sense to combine different representation formats; according to Ainsworth (2006), this can support learners in processing and implementing the feedback and improve their understanding of complementing, constraining, or relating different representations to each other (see also Scaife & Rogers, 1996). Using multimedia or a simple representation format can reduce cognitive load (cf. modality effect; Moreno & Mayer, 2002), but a combination of different representation formats can also lead to no additional learning gain. For instance, Butler (1987) showed that students who received content-related text-based feedback showed more improvement in learning than did students who received only numerical feedback (i.e., grades) or a combination of grades and text-based feedback, probably because the latter focused more on the grade than on the detailed text-based feedback (see also Wiliam, 2007).

The *elaboration level* of feedback refers to the type of feedback that is provided, which depends on the goal and the targeted function of the feedback (Mertens et al., 2022; Panadero & Lipnevich, 2022). For instance, feedback includes less information when the function is verification or communicating an error and the correct response than when the goal is to support students in their self-regulated learning and explain how to apply the appropriate solution strategy. Previous research has indicated that feedback has a differential impact on learning, depending on the level of elaborated of the information (Butler & Winne, 1995; Hattie & Timperley, 2007; Kuklick & Lindner, 2023; Mertens et al., 2022; Shute, 2008).

Narciss (2012) distinguished between different types of feedback related to outcomes, such as knowledge of result, knowledge of correct response, or answer-until-correct feedback. These outcome-related feedback types are also known as corrective feedback (as noted by Fyfe & Rittle-Johnson, 2016; Shute, 2008) or lower-order feedback (in studies like Kuklick & Lindner, 2023; Mertens et al., 2022). The function of these feedback types is to elicit correct responses. However, improving incorrect responses often requires additional information. Therefore, more elaborated feedback types be employed, which offer more substantial feedback content (as discussed by Kuklick & Lindner, 2023; Mertens et al., 2022; Narciss, 2012; Shute, 2008). The function of elaborated feedback is to offer additional guidance and to assist students in developing strategies and consolidating their knowledge so that they may close knowledge gaps and be able to solve future tasks correctly (Brooks et al., 2019; Hattie & Timperley, 2007; Mertens et al., 2022; Shute, 2008).

The *level of complexity* of the task (designated focus of feedback by Patchan et al., 2016) is related to the topic of the feedback information, the implementation processes the task requires, and which mechanisms are activated when processing the feedback information (e.g., decoding, analyzing, interpreting), but it also refers to the type of knowledge that is necessary to solve the task. Generally, independent of the subject domain it can be categorized as a lower level of complexity or a higher level of complexity. In the domain of writing, for example, a lower level of complexity refers to errors of grammar, word length or spelling (see also lower-order learning outcome or lower order of text quality; Mertens et al., 2022; Patchan et al., 2016; Strobl et al., 2019). The text would still be comprehensible, even if there were errors at the lower level of text quality. Accordingly, feedback on a lower level would not be as complex as that on a higher level, such as feedback regarding the overall structure of the text or global cohesion. To implement such complex higher-level feedback, learners must possess both declarative knowledge and domain-specific procedural knowledge to know what the text should look like and how to realize global cohesion. In

contrast, in the domain of physics, for example, the level of complexity could refer to the level of the transfer the learners have to make. For instance, when learners have already solved practice tasks on a specific topic (e.g., Ohm's law), it might be easier for them to transfer this knowledge to a similar task for which the values differ, but the solving steps are the same (i.e., near transfer; lower level of complexity). However, when the task changes and they should apply Ohm's law again but in a different context, they need far transfer knowledge to think beyond the provided information. Previous research indicated that higher-level feedback has a greater impact on learners' performance (McNamara et al., 2013; Patchan et al., 2016), but it has also been shown that students are more likely to implement feedback at a lower level than at a higher level because of the greater complexity of the high-level feedback (Patchan et al., 2016; Strobl et al., 2019).

1.3.3.2. Do Learner-Related Factors make the Difference?

Prior knowledge has been shown to have a crucial impact on students' learning and in the context of feedback (Bryson et al., 1991; Flower et al., 1992; Fyfe & Rittle-Johnson, 2016; Kellogg, 2008; Krause et al., 2009; Mory, 2004; Narciss & Huth, 2004; Patchan et al., 2016). Previous research has suggested that students with lower prior knowledge benefit more from feedback than do students with higher prior knowledge (Allen et al., 2016; Fyfe & Rittle-Johnson, 2016; Krause et al., 2009; Strobl et al., 2019). Shirah and Sidney (2023) provided initial evidence suggesting that computer-based feedback was only effective when the induced prior knowledge was irrelevant to the topic of the targeted task (see also Fyfe & Rittle-Johnson, 2016, which also showed the non-additive effects of inducing topic-related prior knowledge before providing feedback). To this end, more differentiated research is needed on how prior knowledge affects computer-based feedback on learning.

Cognitive capacity is needed to understand the provided information, to decode it, to restructure the information in memory, and to analyze and interpret it (Butler & Winne,

1995; Hattie & Timperley, 2007; Panadero & Lipnevich, 2022). Based on the principles of cognitive load theory (Sweller, 2004; Sweller et al., 1998), it can be assumed that, depending on the amount and presentation of information, the perceived difficulty can be determined, and the mental effort will increase.

Metacognitive processes such as judgment of learning, students' self-assessment of their performance, self-regulation, and monitoring (Labuhn et al., 2010; Zimmerman, 2000), may also play a role in processing the feedback, as previous research has shown that students frequently overestimate their performance and knowledge (Baars et al., 2013; Finn & Metcalfe, 2014). Thus, Fyfe and Rittle-Johnson (2016, 2017) and Narciss (2013) assumed that when prior knowledge is induced first, students will no longer focus on the feedback because they are already feeling competent. This inadequate self-assessment can have a potential impact on learners' self-regulation during feedback processing and can reduce performance.

Affective-motivational processes such as engagement, pleasantness, valence, arousal, or intrinsic motivation could also influence learning (Hattie & Timperley, 2007; Mory, 2004; Shute, 2008). Previous research has indicated that such affective-motivational factors seem to be crucial because they can both facilitate and inhibit students' cognitive processes, intrinsic motivation, and competence beliefs, which consequently affects their performance (Finn, 2015; Fong et al., 2019; Kluger & DeNisi, 1996; Kuklick & Lindner, 2021; Mertens et al., 2022). Valence (i.e., whether learners receive positive or negative feedback) can additionally strengthen or reduce the effects of affective-motivational processes (Burgers et al., 2015; Efklides & Dina, 2004; Fong et al., 2019; Grundmann et al., 2021; Kluger & DeNisi, 1996; Kuklick & Lindner, 2021, 2023). Thus, previous researchers have argued that future research is needed to investigate the relationship among affective components, feedback, and learner outcome performance (e.g., Goleman, 1995; Picard et al., 2004; Shute, 2008).

1.3.3.3. Do Context-Related Factors make the Difference?

The context of the feedback situation plays a crucial role in the effectiveness of feedback (Hattie & Clarke, 2019; Narciss & Huth, 2004). Nowadays, the understanding that feedback is subject-specific is widespread. However, there are other possible contextual influencing factors that have received less attention.

The *orchestration of computer-based feedback* refers to the instructional context and describes whether feedback is provided as a single intervention or combined with other instructional strategies (e.g., strategy instruction). Previous research on the *orchestration of feedback* has provided initial evidence across disciplines that feedback alone is more effective in fostering learning than feedback combined with strategy instruction (Fyfe & Rittle-Johnson, 2016; Shirah & Sidney, 2023; Wischgoll, 2017). However, Graham et al. (2015) and Salden et al. (2010) reported contrasting findings regarding whether the combination of feedback improved learning. To shed light on and explain these heterogeneous results, a stronger empirical investigation of the feedback linked to strategy instruction is needed. In this dissertation, the focus was on investigating the orchestration of computer-based feedback combined with prior strategy instruction.

The *sequence* in which feedback is provided when it is combined with instruction is included separately as a further factor related to the instructional context. Sequence was added as a separate factor because it does not describe the orchestration of prior instruction with additional feedback but rather when the computer-based feedback is provided (e.g., at the beginning of the learning process or in between the learning process). The regular sequence starts the learning process with a learning phase (including instruction or worked examples), followed by a practice phase that includes feedback (I-PS [instruction followed by problem solving]; e.g., Kant et al., 2017; van Gog et al., 2011). In contrast, the learning process can also start with a practice phase including feedback followed by a subsequent learning phase (PS-I [problem solving prior instruction]; e.g., Kapur, 2012; Loibl et al.,

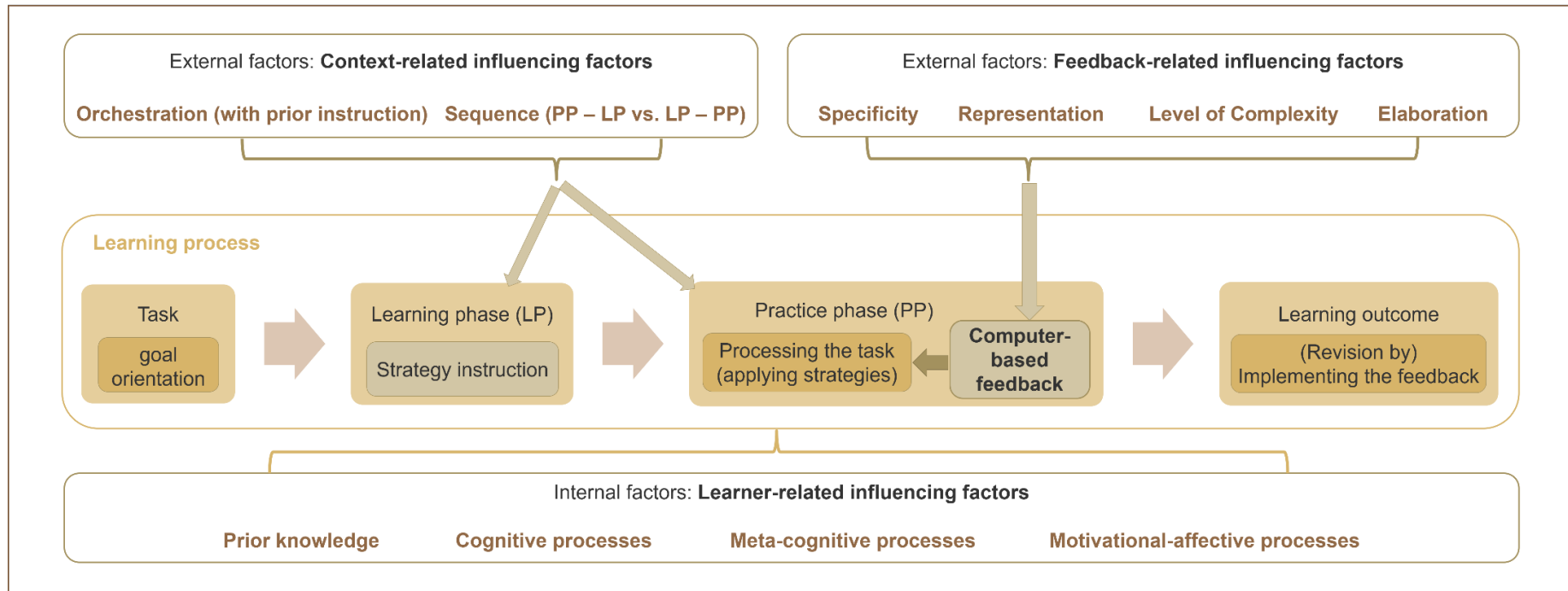
2017). Previous research on the sequence of feedback in an instructional context has produced contradictory findings regarding the sequence in which feedback should be provided when it is combined with instruction. On the one hand, research in the field of PS-I (see Kapur, 2012; Loibl et al., 2017; Sinha & Kapur, 2021) has indicated that providing students with a practice phase followed by instruction yields higher conceptual and transfer knowledge than does providing instruction with a subsequent practice phase (Sinha & Kapur, 2021). On the other hand, previous research has shown that when instruction is followed by a practice phase, students profit more than when the learning phase is preponed (Kant et al., 2017; van Gog et al., 2011). Despite this, previous research on sequence has only examined the difference between PS-I and I-PS and has not specifically considered feedback in the practice phase. Thus, it is still unclear if feedback supports learning more when it is provided at the beginning of a teaching session or after instruction.

1.4 The CAFCALI Model

In order to organize the identified potential influencing factors of the effect of computer-based feedback on learning (see Chapter 1.3.3), the CAFCALI (Computer-Assisted Feedback-, Context-, And Learner-related Impact on Learning) model was developed as part of this dissertation (see Figure 2).

Figure 2

The CAFCALI Model



Note. CAFCALI stands for Computer-Assisted Feedback-, Context-, And Learner-related Impact on Learning.

Previous theoretical feedback models include most of the named relevant influencing factors (as described in Chapter 1.3.3 Potential Influencing Factors) related to the implementation of feedback. However, in the MISCA model (Panadero & Lipnevich, 2022), the individual factors were not mentioned directly in the model but grouped together under a category. For example, the MISCA model, includes the category *implementation*, which comprises internal processing and function. Panadero and Lipnevich (2022) explained that this category refers to students' internal processing of feedback information (e.g., decoding, analyzing, interpreting dependent on the function of the feedback), which is related to students' cognitive, metacognitive, and affective-motivational activities. These underlying processes, in turn, differ due to individual differences that were classified under *students' characteristics*. It is therefore not entirely clear from which perspective the factors are viewed, that is, from a feedback perspective (e.g., depending on the function of the feedback) or from a learner perspective (e.g., the underlying processes the feedback initiates). In addition, the learning process itself is not explicitly included.

Narciss and Huth (2004) chose a different approach: dividing their framework into the categories *instruction*, *feedback*, and *learner*. The authors addressed the learning process itself by including the aspects of objectives, tasks, and errors under the component *instruction*, which is related to the learning goal and the necessary skills and types of knowledge. It should be noted that this framework was developed as a guideline for how to design feedback. Thus, these named aspects of instruction must also be interpreted as feedback-related aspects. For instance, whether the feedback explains the source of error, which strategy was incorrect, or repeating the learning goals but not explicitly considering how feedback is implemented in the context of the learning process and the sequence in which it is orchestrated. In this framework, cognitive, metacognitive, and motivational aspects appear in all three categories (i.e., instruction, feedback, learner), but from a different perspective in each case. From an instructional perspective, for example, the authors stated

that cognitive operations and metacognitive skills are relevant to understanding the task itself. In the feedback category, they mentioned the cognitive, metacognitive, and motivational processes that are activated by the feedback, depending on the function of the feedback (as described in the MISCA model). Separately, they noted the influence of feedback on the learner and their performance by including the influences of motivation, prior knowledge, and metacognitive skills at an individual level. In contrast to the MISCA model, in the framework of Narciss and Huth (2004), the *content* of the feedback mainly referred to the level of elaboration; aspects such as specificity or complexity were not specifically mentioned. Likewise, the instructional category mainly referred to the learning objectives, tasks, and errors, and less to how feedback was orchestrated.

In contrast, the ITFL-model by Narciss (2008) takes the learning process into account and considers learner-related (internal) factors as well as (external) factors based in the learning and teaching context. On both sides, the learning process starts with goal orientation and setting as well as individual or technical prerequisites. Building an internal or external representation of the standards, competences, and task requirements allows for identification of the internal and external target values of the desired level of competence. Subsequently, a learning phase with knowledge or skill acquisition takes place, followed by task processing, which refers to the actual value of students' performance. According to Narciss (2008), there are two feedback loops, one internal and one external, which interact with each other. In the internal feedback loop, learners' cognitive, metacognitive, and affective-motivational processes are activated, and learners act accordingly when processing and implementing the feedback to reach the internal targeted value. On the other hand, the model also compares the actual and external targeted values. Based on this evaluation, elaborated and tutorial information can be provided. It is clear that the ITFL-model is very comprehensive and considers the learning process, context-related factors, and learner-

related factors. However, no explicit feedback factors, such as representation or the specificity of the feedback, are included in the model.

The named models refer to some degree to the instructional context of the feedback (*instructional medium* in the ITFL-model by Narciss, 2008; *instruction* as an element of the framework by Narciss & Huth, 2004; and the *context* component in the MISCA model by Panadero & Lipnevich, 2022). However, the orchestration of computer-based feedback, in the sense of its combination with other instructional strategies and how, for example, prior presented instruction may have an impact on the effect of computer-based feedback, is not explicitly mentioned.

Many models, therefore, contain important aspects for testing the influencing factors of computer-based feedback, but to date there exists no model that depicts the relevant feedback-, learner-related and context-related factors (i.e., orchestration of computer-based feedback) directly in the model, nor their relationship to each other and their embedding in the learning process. The CAFCALI model was developed to address this lack and to investigate under which circumstances computer-based feedback fosters learning.

In contrast to most previous feedback models, the CAFCALI model explicitly considers the learning process, which is represented in the center of the model (see Figure 2). The CAFCALI model describes the learning process in four steps (based on Bangert-Drowns et al., 1991; Narciss, 2008): First, students receive a task and become familiar with the goal of the learning unit. Second, in the learning phase, students receive input via a question or a provision of a strategy instruction. Third, in the practice phase, students respond to the question or try to apply the strategies they were introduced to by the prior instruction. Last, as a fourth step, learners adjust their strategies to solve the task, close knowledge gaps, or revise their responses by implementing the information they received through the computer-based feedback (cf. mindfulness model by Bangert-Drowns et al., 1991).

In addition, the CAFCALI model takes the required multidimensional perspective by considering the theoretically derived potential influencing factors of computer-based feedback that are relevant for implementation, orchestration, and processing of computer-based feedback to support the learning process across subject domains. The CAFCALI model differs regarding context-related factors (i.e., orchestration with prior strategy instruction, sequence) and feedback-related factors (i.e., specificity, representation, elaboration, level of complexity). The internal factors refer to the learner-related factors (i.e., prior knowledge, cognitive, metacognitive, and affective-motivational processes). Thereby, it follows the distinction drawn by Narciss (2008) between external and internal factors. The various influencing factors are shown around the learning process (see Figure 2). Arrows illustrate references to the learning process that are relevant for investigating the effect of computer-based feedback.

As learners-related factors play a role in multiple stages of the learning process, they are presented at the bottom as the foundation of the CAFCALI model (see Figure 2). Before learners begin with a learning process, they already possess a current state of prior knowledge that act as a base for the processing of new information (Bangert-Drowns et al., 1991; Shute, 2008). In addition, it might play a role in how motivated learners are to learn something new or to process the task when the learning process starts (Jiang et al., 2018; Ryan & Deci, 2000). Their motivational state might depend on their interest in the topic or their affective mood (Ainley et al., 2005; Pekrun, 2005). When learners receive a task, a certain level of cognitive capacity is necessary just to understand it (Atkinson & Shiffrin, 1968; Renkl, 2009). Likewise, in the learning phase, students' cognitive capacities are used up because prior knowledge (in the form of information located in memory) is activated. Depending on their current state of prior knowledge, learners monitor their learning process and might perform assessments of their own performance (cf. internal loop; Narciss, 2008). So, in the practice phase different cognitive, metacognitive, and affective-motivational

mechanisms also come into play in processing the provided information. Based on these mechanisms, the learners might have any kind of expectation of the feedback (Eccles & Wigfield, 2020; Kuklick & Lindner, 2021; Narciss, 2008). When students receive feedback after practicing, evaluating the feedback information can also activate different cognitive, metacognitive, and affective-motivational processes. Processing feedback information requires cognitive capacity and can lead to cognitive overload if the same information was provided in a prior instruction (Sweller et al., 1998). In addition, learners might be disappointed (when they overestimated their performance) or happy (when they underestimated their performance or assessed their performance correctly), depending on their metacognitive judgments. This could lead to an increase or decrease in motivation or learners' arousal. The amount and the direction (e.g., positive increase of motivation or negative increase of extraneous cognitive load) can influence the uptake and implementation of feedback. Since the internal learner-related influencing factors are activated at different stages of the learning process, there is only one arrow from all learner-related factors to the learning process as a whole. In this dissertation, the focus is on which underlying processes are triggered by computer-based feedback and on whether conscious prior knowledge induction influences the effect of computer-based feedback.

Opposite the internal factors are the external factors presented (at the top of Figure 2). These are divided into context-related and feedback-related factors. Among the context-related factors, orchestration refers to whether the practice phase with the feedback is preceded by a learning phase (in which, for example, a strategy instruction is provided) or whether the feedback is used as the only support strategy in the learning process. The sequence refers to the point in time when the feedback is provided, either after a learning phase, as in most practical cases, or at the beginning of the learning process (i.e. a practical phase followed by a learning phase). The use of two arrows indicates that both context-related factors refer to both the learning and the practice phases (see Figure 2). The feedback-

related factors refer to the design of the feedback information (i.e., specificity or elaboration, representation, and the level or complexity), which is why one arrow represents the direct relationship between all feedback-related factors and the feedback element of the learning process.

1.5 Goal and Research Questions

Learning can be a challenging process, especially in individual learning phases, but feedback is considered an effective instructional strategy to support learning. Feedback is a common term in modern teaching and learning research, but there are still many different conceptualizations. At the same time, there exists a common ground in the educational field, as specific core elements of the feedback process appear in almost all theoretical feedback models. In this dissertation, feedback in general was defined as post-response information following performance that is provided by an external agent (i.e., computer) on aspects of a learners' current performance and is intended to help the learner close the gap between the current performance and the target performance based on success criteria (Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Lipnevich & Panadero, 2021; Narciss, 2008, 2013; Shute, 2008). As computer-based feedback is regarded as a helpful supplement in teaching and individual learning situations, both in the classroom and in online environments (e.g., Bangert-Drowns et al., 1991; Mertens et al., 2022; van der Kleij et al., 2015), several feedback technologies and platforms or adaptive systems to provide teacher-mediated automatic feedback have been developed over the years.

Although there exist some meta-analyses focusing mostly on the type of feedback or on the impact of students' prior knowledge, there is still a lack of meta-analytic evidence on the use of computer-based feedback to foster learning. Furthermore, research on the orchestration of computer-based feedback with further instructional strategies commonly used in practice (e.g., prior strategy instruction) is altogether missing (see Fyfe & Rittle-

Johnson, 2016; Shirah & Sidney, 2023 for expectations). The few existing studies that examine at the orchestration of feedback focus mainly on the impact of prior knowledge. Other potential student-related influencing factors (e.g., cognitive, metacognitive, or affective-motivational processes), as well as feedback-related influencing factors, were not examined.

Previous researchers assumed that one potential explanation for the inconsistent findings regarding the effect of computer-based feedback might be the mediating impact of different boundary conditions. Related evidence (e.g., Burkhart et al., 2020; Kuklick & Lindner, 2023; Lachner et al., 2017a, 2017b; Mertens et al., 2022; Narciss, 2008, 2013; Patchan et al., 2016; Strobl et al., 2019) indicated that there are various influencing factors that must be considered during both provision and implementation of feedback to support learning. This dissertation developed the theory driven CAFCALI model based on previous feedback models. The CAFCALI model describes how feedback is embedded in the learning process and how factors related to the feedback message (i.e., content, specificity, representation, elaboration), the learner (i.e., prior knowledge, cognition, meta-cognition, affection/motivation), and the context (i.e., orchestration, sequence) can have an impact on the degree to which of computer-based feedback fosters learning. In previous research, only some of these potential influencing factors were investigated. Thus, it is still an open question under which conditions computer-based feedback is effective at increasing students' learning gain (e.g., by improving students' writing quality or students' performance in problem solving tasks).

Therefore, there is a need to take a multidimensional perspective considering the “situational and individual characteristics of the instructional context and learner [...] along with the nature and quality of a feedback message” (Shute, 2008, p. 176; see also Strobl et al., 2019). In particular, it is important to examine the facets of feedback, learners, and

instructional context, both separately and interactively (Narciss & Huth, 2004; Shute, 2008; Strobl et al., 2019).

To address these needs, the overall aim of this dissertation was to investigate under which boundary conditions computer-based feedback fosters learning in different subject domains. This led to the overarching research question (RQ):

What are the factors that significantly influence the effect of computer-based feedback on learning?

1.6 Overview of the Studies of this Dissertation

To answer the overarching RQ meaningfully, this dissertation comprises one theoretical paper and two empirical papers containing one meta-analysis and three online experiments that investigated the effects of computer-based feedback on learning in different subject domains (i.e., explanatory writing; problem solving in physics) as well as explore and manipulate several potential influencing factors (see Figure 3).

The first paper, “Feedback – Ja, klar?!” addressed the RQ by describing the potential of computer-based feedback tools to support the learning process of students who are developing their exploratory writing skills. Thereby, I focused on the feedback-related boundary conditions (i.e., specificity, representation, complexity of content; see Figure 3).

Based on this conceptualization of feedback, I conducted a meta-analysis (Paper 2) that addressed the RQ by examining the overall effect of computer-based feedback technologies regarding supporting students’ expository writing, as well as explore the impact of potential feedback- and learner-related influencing factors (see Figure 3) that were derived from previous research on the implementation of computer-based feedback and identified as possibly relevant in Paper 1.

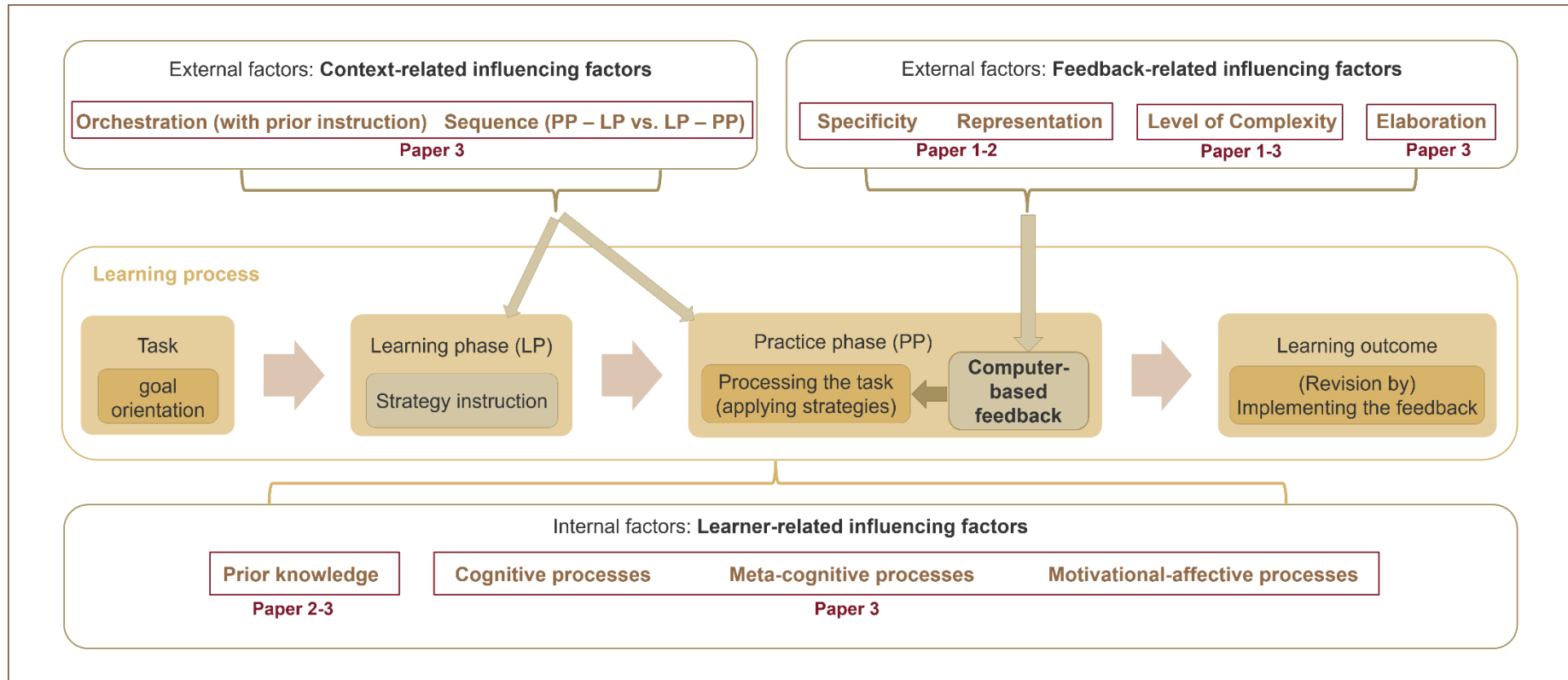
Paper 3 presented three online experiments in which different experimental manipulations were realized. In Experiments 1 and 2, I focused on the effectivity of corrective (Experiment

Overview of the Studies of this Dissertation

1) and elaborated (Experiment 2) feedback on students' physics learning (i.e., problem solving in the context of Ohm's law). Addressing the RQ, I focused on the orchestration of computer-based feedback with strategy instruction in all three online experiments. Furthermore, I investigated different feedback-related factors (i.e., elaboration, complexity of content) and, again, learners' prior knowledge as well as further learner-related influencing factors (i.e., cognitive, metacognitive, and affective-motivational aspects). In addition to orchestration, Experiment 3 also investigated another context-related factor: sequence (see Figure 3).

Figure 3

Assignment of the Studies into the CAFCALI Model



More explicitly, Paper 1 aimed to conceptualize the potential of digital media, with a focus on computer-based feedback. Based on a brief conceptual introduction to the acquisition of writing skills that included reference to cognitive psychological processing models, the paper discussed research-based potentials and both the risks of computer-based feedback systems in writing acquisition and their use for fostering writing skills. Thereby, potential influencing factors were identified, after which the paper presented evidence-based tools for writing feedback. Finally, it concluded with a discussion of potential further developments and research desiderata.

Paper 2 aimed to address these research desiderata. The aim of the meta-analysis was to examine the effectiveness of formative computer-based feedback that is generated automatically by a system to foster students' writing quality in their first language, as well as to analyze potential influencing factors. Thereby, I tested the potential influencing factors with a focus on those related to feedback (named in Paper 1). Only studies with a pretest-posttest design ($k = 24$ effects) or a pretest-posttest-control group design ($k = 15$ effects) were included.

In Paper 3, the overarching aim was to explore the potential synergy- or interaction effects of combining computer-based feedback with strategy instruction on student learning in the domain of physics, as well as to rigorously examine the generalizability of these potential combination effects. Therefore, I conducted three online experiments ($N_1 = 437$, $N_2 = 310$, $N_3 = 166$). In order to identify internal determining factors, I examined the underlying cognitive (i.e., mental effort, subjective difficulty), metacognitive (i.e., monitoring accuracy) and affective-motivational (i.e., arousal) processes to explain additive versus reducing effects of the combination of strategy instruction and computer-based feedback. Regarding the external feedback-related factors, I varied the type of format by applying corrective feedback in Experiment 1 and elaborated feedback in Experiment 2 with the aim of testing the robustness of the effects I obtained in Experiment 1. Regarding the external context-

related factors, I manipulated the sequence, (i.e., whether strategy instruction was followed by feedback or vice versa) and investigated whether feedback was processed differently depending on the timing.

2 FEEDBACK – JA, KLAR?!

Wagner, S. & Lachner, A. (2021). Feedback – Ja, klar?! Digitale Medien zur Förderung von Schreibkompetenzen. *leseforum.ch*, 3, 1–16.

This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. The Version of Record of this manuscript has been published and is available in the journal *leseforum.ch* in 2021.

Abstract

Damit Schreibende verständliche und elaborierte Texte verfassen können, müssen sie distinkte Schreibstrategien entwickeln, welche oft hohe Anforderungen an diese stellen. Im vorliegenden Beitrag wird vor diesem Hintergrund das Potenzial digitaler Medien zur Förderung von Schreibstrategien diskutiert. Besonderer Fokus liegt hierbei auf der Unterstützung von Revisionsprozessen mittels computerbasierten Feedbacks. Hierzu wird zunächst eine forschungsbasierte Taxonomie vorgeschlagen, welche auf kognitionspsychologischen Theorien des Schreiberwerbs beruht, anhand derer digitale Feedbacktechnologien klassifiziert werden. Basierend auf dieser Taxonomie werden exemplarisch drei Feedbacktechnologien vorgestellt. Potenzielle Entwicklungs- und Forschungsdesiderate werden abschließend diskutiert.

Schlüsselwörter

Schreiberwerb, Bildungstechnologie, computerbasiertes Feedback, digitale Medien, Schreibförderung

2.1 Einleitung

Um elaborete und verständliche Texte verfassen zu können, sollen Kinder und Jugendliche nachhaltige Schreibkompetenzen entwickeln, so dass sie Schreibprodukte (Textentwürfe und fertige Texte) verfassen können, die komplexe und vielschichtige Informationen beinhalten. Idealerweise sollten diese Texte so gestaltet sein, dass sie adressatengerecht aufbereitet sind (Graham & Perin, 2007; Lachner et al., 2017a). Schreiben stellt jedoch hohe Anforderungen; daher brauchen Kinder und Jugendliche aber auch erwachsene Lernende instruktionale Unterstützung, um adäquate Schreibkompetenzen zu entwickeln (Graham & Perin, 2007). Digitale Medien werden hierbei als zentrale Möglichkeit sowohl in den Bildungswissenschaften als auch den Fachdidaktiken diskutiert, um adaptive Unterstützungsmöglichkeiten zu bieten (Lachner & Scheiter, 2020). Das Hauptaugenmerk medienbezogener Forschung lag bisher stark auf der Nutzung digitaler Medien in den MINT-Fächern (siehe Yeung et al., 2021 als Überblick). In diesem Beitrag wollen wir daher Herausforderungen und Potenziale bezüglich der Nutzung digitaler Medien für den Schreiberwerb als zentrale Facette des Sprachenunterrichts (Erstspracherwerb) skizzieren. Nach einer kurzen konzeptuellen Einführung zum Erwerb von Schreibkompetenzen unter Bezugnahme kognitionspsychologischer Verarbeitungsmodelle werden forschungsbasierte Potenziale und Risiken digitaler Medien beim Schreiberwerb diskutiert. Hierbei wird der Fokus auf computerbasierte Rückmeldesysteme als Methode zur Förderung von Schreibkompetenzen mit digitalen Medien gesetzt. Anschließend werden evidenzbasierte Tools zur Rückmeldung beim Schreiben vorgestellt. Auf Basis dieser Vorstellung werden potenzielle Weiterentwicklungen sowie Forschungsdesiderate diskutiert.

In diesem Beitrag nehmen wir nicht Bezug auf eine besondere Zielgruppe, da die Tools in unterschiedlichen Bildungsbereichen (bspw. verschiedene Schultypen oder Universitäten)

eingesetzt werden und wir hier allgemeine Aspekte ansprechen, die sich auch auf das Lernen Erwachsener übertragen lassen können.

2.2 Die Entwicklung von Schreibkompetenzen

Im folgenden Beitrag wird eine psychologische Perspektive auf Schreibkompetenzen eingenommen. Daher werden unter Schreibkompetenzen die bei Lernenden verfügbaren kognitiven Ressourcen (d. h. Fertigkeiten und Wissen) verstanden, um bestimmte Schreibprobleme zu lösen sowie die damit verbundenen motivationalen, volitionalen und sozialen Ressourcen, um die Problemlösungen in verschiedenen Situationen erfolgreich zu nutzen (Becker-Mrotzek & Schindler, 2007; Ossner, 2006).

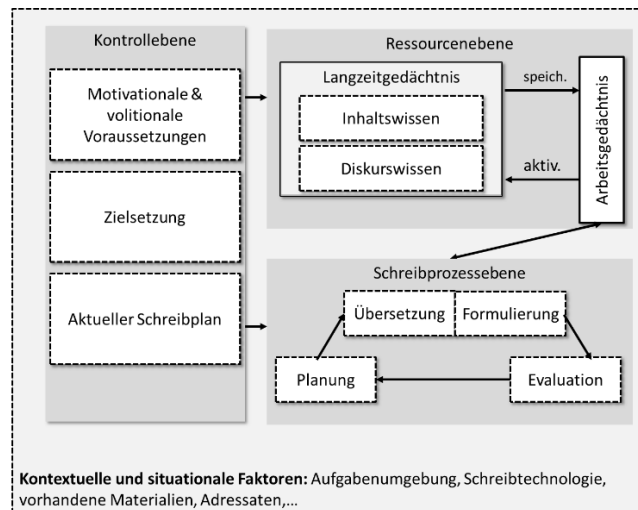
Aus kognitionspsychologischer Perspektive unterscheiden sich Schreibexpert*innen und Noviz*innen jedoch nicht nur hinsichtlich ihrer verfügbaren Ressourcen, sondern ebenfalls dahingehend, welche Schreibstrategien sie (im Schreibprozess) berücksichtigen. Nach dem Modell von Hayes (2012) lassen sich auf Prozessebene idealerweise vier zyklisch aufeinander bezogene Schreibprozesse unterscheiden (siehe Figure 4): Zunächst entwerfen Schreibende einen Plan ihres Textes, welcher anschließend in ein konzeptuelles Modell des Textes übersetzt wird. Dieses konzeptuelle Modell bildet die Grundlage für eine erste Formulierung eines Entwurfs ihres Textes. Anschließend wird dieser Entwurf evaluiert. Diese Evaluation kann anschließend in eine weitere Revision des Textes münden. Um diese Schreibprozesse effektiv zu bewerkstelligen, ist auf kognitiver Ebene sowohl Inhaltswissen (d. h. inhaltliches Wissen über das Thema), als auch Diskurswissen (d. h. rhetorisches Wissen über Schreibstrategien, um Wissen verständlich und kohäsiv zu präsentieren) erforderlich. Beide Wissensarten müssen dabei integriert werden (Becker-Mrotzek et al., 2014; Bereiter & Scardamalia, 1987). Neben kognitiven Faktoren hängt die Qualität der jeweiligen Schreibprodukte und das kognitive Engagement der Schreibenden von den motivational-volitionalen Voraussetzungen der Schreibenden ab (Azevedo, 2015;

Die Entwicklung von Schreibkompetenzen

Duijnhouwer et al., 2012), sowie von kontextuellen Faktoren, die beispielsweise unterschiedliche Affekte während des Schreibens evozieren (D’Mello & Graesser, 2012).

Figure 4

Kognitive und metakognitive Schreibprozesse, adaptiert nach Hayes (2012)



Gleichzeitig verdeutlichen empirische Befunde, dass das Schreiben und die damit einhergehenden kognitiven, motivationalen und affektiven Schreibprozesse hohe Anforderungen an Schreibnoviz*innen stellen, denn Noviz*innen verfügen noch nicht in vollem Umfang über solche Schreibstrategien. So zeigen empirische Untersuchungen, dass Noviz*innen weniger Zeit in die Planung investieren (Bryson et al., 1991; Flower et al., 1992). Zudem tendieren Noviz*innen zu einem sogenannten Knowledge-Telling Bias, bei dem sie alle verfügbare Information aufschreiben, ohne die Zielgruppe und das Textmuster zu berücksichtigen (Kellogg, 2008). Diese Effekte verschärfen sich insbesondere dadurch, dass Noviz*innen dazu tendieren, wenige Überarbeitungen zu realisieren und, wenn überhaupt, nur oberflächliche Überarbeitungen, wie orthografische oder syntaktische Änderungen vornehmen, die weniger die eigentliche Qualität des Textes verbessern (Chanquoy, 2009; Sturm, 2016; van Steendam et al., 2014). Vor diesem Hintergrund ist es wichtig, didaktische Verfahren zu identifizieren, die insbesondere die Überarbeitung beim

Schreiben fokussieren. Auf dieser Grundlage wird Feedback als zentrale Maßnahme zur Förderung von Revisionen diskutiert (Cho & MacArthur, 2010; Graham et al., 2015; Kellogg & Whiteford, 2009; Lachner et al., 2017a).

2.3 Feedback zur Unterstützung von Revisionsprozessen

Feedback wird gemeinhin als Information bzgl. verschiedener Aspekte der Leistung von Lernenden (in diesem Fall in Bezug auf die Schreibqualität) verstanden und kann sowohl zur Bewertung als auch zur Diagnostik oder Evaluation von Leistungen anhand verschiedener Erfolgskriterien dienen (Hattie & Timperley, 2007; Kluger & DeNisi, 1996). Hierbei wird formatives und summatives Feedback unterschieden.

Formatives Feedback wird (mehrmals) während des Schreibprozesses bereitgestellt und kann sich auf den Textentwurf und auf die Entwicklung bzw. Verbesserung von Schreibstrategien beziehen. Summatives Feedback wird erst am Ende des Schreibprozesses auf das fertige Schreibprodukt gegeben. Das Feedback kann sowohl von Lehrpersonen als auch von Peers oder systemgeneriert zur Verfügung gestellt werden (MacArthur, 2015). Im folgenden Beitrag wollen wir uns auf computerbasiertes formatives Feedback fokussieren, da dieses leicht in individuellen Lernsettings eingesetzt werden kann. Nichtsdestotrotz wollen wir hier auf die hohe Effektivität von Peer-Feedback verweisen, da dieses nicht nur Lernende unterstützen kann, die Feedback erhalten (Cho & MacArthur, 2010), sondern insbesondere auch diejenigen, die Feedback geben (Cho & MacArthur, 2011). Idealerweise können solche Peer-Feedbackansätze auch auf kollaborative Schreibsettings erweitert werden.

Ziel von Feedback ist es, Lernende dabei zu unterstützen, die Diskrepanz zwischen der aktuellen Leistung (z. B. die Qualität des aktuellen Textentwurfs) und der zu erzielenden Performanz (Ziele des Schreibprodukts) zu ermitteln und zu reduzieren (Stevenson & Phakiti, 2014; Winne & Butler, 1994). Feedback resultiert dabei als potenzielle Konsequenz

Computerbasiertes Feedback mit digitalen Medien – eine Methode zur Förderung der Schreibkompetenz von Lernenden von Verhalten der Lernenden und kann als indirekte Fördermaßnahme im Sinne von Scaffolding angesehen werden (Hattie & Timperley, 2007).

Jüngere Metaanalysen stützen die angenommenen Potenziale formativen Feedbacks zur Entwicklung der Schreibkompetenz. Beispielsweise fanden Graham et al. (2011) einen mittleren bis großen Effekt von Feedback auf Schreibqualität von $d = 0.77$ (im Vergleich zu einer anderen Treatmentgruppe oder einer Gruppe, die kein Feedback erhielt), basierend auf 16 empirischen Studien. Obwohl Feedback als wichtige Methode zur Förderung von Schreibkompetenz angesehen wird, spielt es bislang in (Hoch-)schulen in der Regel jedoch eine eher untergeordnete Rolle. Dies mag darauf zurückzuführen sein, dass individuelles und formatives, also mehrmaliges, Feedback zeitintensiv und aufwendig ist, und daher selten in angewandten Kontexten realisiert werden kann. Alternativ wird hier die Rolle digitaler Medien, insbesondere in individuellen Lernphasen (z. B. während der Erledigung von Hausaufgaben), diskutiert. Digitale Medien ermöglichen, computervermittelte Rückmeldung zu geben (bspw. indem Lehrpersonen oder Peers mithilfe des computerbasierten Tools Feedback zu einem Schreibprodukt geben können) oder das Feedback komplett von einem intelligenten System bereitstellen zu lassen, welches Methoden der künstlichen Intelligenz (KI) und/oder Natural-Language-Processing-Technologien (NLP) nutzt, um hierarchieniedrigere Überarbeitungen des Textes bei Noviz*innen anzuregen (Lachner et al., 2017a; Roscoe & McNamara, 2013; Wilson et al., 2017).

2.4 Computerbasiertes Feedback mit digitalen Medien – eine Methode zur Förderung der Schreibkompetenz von Lernenden

Computerbasierte Feedbacksysteme sind wichtige Anwendungen digitaler Medien, um den Schreiberwerb von Lernenden zu fördern. Digitale Medien sind insbesondere geeignet, um Rückmeldung über operationalisierbare Merkmale von Schreibprodukten ad-hoc zu geben (z. B. Orthografie, Kohäsion). Aus technologischer Perspektive scheint

Computerbasiertes Feedback mit digitalen Medien – eine Methode zur Förderung der Schreibkompetenz von Lernenden

computerbasiertes Feedback sogar einen Vorteil gegenüber menschlichen Rater*innen (Bewerter*innen) zu bieten. Beispielsweise zeigen empirische Befunde, dass Feedbacksysteme in der Lage sind, bei operationalisierbaren Merkmalen akkuratere Rückmeldungen zu geben als menschliche Rater (siehe z. B. Burkhart et al., 2020). Dagegen sind computerbasierte Systeme weniger in der Lage, Rückmeldung hinsichtlich tiefergehender Merkmale von Texten zu geben (z. B. inhaltliche Adäquatheit, Passung zum Genre, siehe Crossley & McNamara, 2011). Allerdings zeigen sich hier oft auch zwischen menschlichen Ratern große Unterschiede.

Trotz dieser Potenziale bescheinigen metaanalytische Studien über die reine Nutzung computerbasierten Feedbacks keine hohe Effektivität. So fanden Graham et al. (2011) lediglich kleine, nicht signifikante Effekte von computerbasiertem Feedback im Vergleich zu Business-As-Usual-Interventionen in ihrer Metaanalyse. Diese Ergebnisse beruhten allerdings auf der Zusammenfassung von vier sehr unterschiedlichen empirischen Studien mit heterogenen Ergebnissen, was die statistische Aussagekraft der Ergebnisse insbesondere bezüglich der Signifikanz des Effekts einschränkt. Außerdem verdeutlichen die heterogenen Befunde, dass computerbasiertes Feedback lediglich ein instruktionales Angebot darstellt, welches von Lernenden genutzt wird. Neben Merkmalen des Feedbacks kommt es deshalb insbesondere auf die kognitiven und motivationalen Voraussetzungen der Lernenden an, wie grundlegende Forschung zu (computerbasiertem) Feedback (fachübergreifend) nahelegt. Beispielsweise untersuchten Fyfe und Rittle-Johnson (2016) den Einfluss von Vorwissen und formativem Feedback auf das Problemlösen von Schüler*innen im Bereich elementarer Mathematik. Die Autorinnen konnten zeigen, dass lediglich Schüler*innen mit weniger Vorwissen von dem Feedback profitierten. Für Schüler*innen, die über viel Vorwissen hatte die Zugabe von Feedback sogar negative Effekte auf den Lernerfolg. Im Bereich akademischen Schreibens untersuchte Wischgoll (2017) den Zusammenhang von Feedback und Strategieinstruktion, indem eine Gruppe keine Instruktion erhielt und sie bei zwei

Computerbasiertes Feedback mit digitalen Medien – eine Methode zur Förderung der Schreibkompetenz von Lernenden
weiteren Gruppen Vorwissen durch die Instruktion einer Zusammenfassungsstrategie oder einer Strukturierungsstrategie induzierte. Zusätzlich erhielt eine Hälfte der Studierenden Feedback in Form von standardisierten Prompts, während die andere Hälfte kein Feedback erhielt.

Ähnlich wie Fyfe und Rittle-Johnson (2016) fand Wischgoll (2017), dass Studierende mit induziertem Wissen nicht vom Feedback profitierten. Die Gruppe von Studierenden, die die Kombination aus Kontrollinstruktion und standardisiertem Feedback erhielt, übertraf Studierende, die explizite Strukturierungsinstruktion und Feedback bekamen hinsichtlich der Textqualität bei akademischen Schreibaufgaben. Neben dem Vorwissen scheint computerbasiertes Feedback auch von den subjektiven Überzeugungen moderiert zu sein. Im Kontext von Textverstehen variierten Golke et al. (2015) die Art des Feedbacks (computer-mediert versus personen-mediert) bei Konstanzhaltung der gegebenen Feedbackinformation und verglichen dieses mit einer Kontrollgruppe, die kein Feedback erhielt. Golke et al. (2015) konnten zeigen, dass Lernende, die personen-mediertes Feedback erhielten, die beiden anderen Bedingungen übertrafen. Entgegen der Erwartung zeigten sich keine Unterschiede zwischen der computer-medierten Feedbackbedingung und der Kontrollgruppe. Diese Befunde verdeutlichen, dass die Effektivität von Feedback zum einen von den individuellen Eigenschaften und Überzeugungen der Schreibenden determiniert wird (siehe auch Kluger & DeNisi, 1996; Patchan et al., 2016), wobei diese Forschung größtenteils noch im Kontext Schreiben repliziert werden sollte. Zum anderen wird deutlich, dass neben dieser inter-individuellen Perspektive auch computerbasiertes Feedback bestimmte Merkmale aufweisen sollte (siehe Kapitel 2.5 Dimensionen computerbasierter Feedbacks), um distinkte Schreibstrategien zu fördern. Neuere Studien berichten durchaus von lernförderlichen Ansätzen, anhand derer die Schreibkompetenz von Lernenden durch computerbasiertes Feedback gesteigert werden kann (z. B. Graham et al.,

2015; Lachner et al., 2017a; McNamara et al., 2015; Roscoe & McNamara, 2013).
Potenzielle Ursachen auf Feedbackebene werden im folgenden Kapitel diskutiert.

2.5 Dimensionen computerbasierter Feedbacks

Computerbasierte Feedbacksysteme können im Hinblick auf drei Hauptdimensionen unterschieden werden: 1) die Dimension der Textqualität (Worauf wird Feedback gegeben?), 2) die Dimension der Feedbackrepräsentationsform (Wie wird das Feedback präsentiert?) und 3) die Spezifität des Feedbacks (Auf welche Art und Weise wird Feedback gegeben?).

2.5.1 Dimension der Textqualität

Die Qualität eines Textes kann grob in eine hierarchieniedrigere und eine hierarchiehöhere Ebene eingeteilt werden (Chanquoy, 2009; Patchan et al., 2016; Strobl et al., 2019). Rückmeldung auf hierarchieniedrigerer Textebene fokussiert Oberflächenmerkmale und betrifft beispielsweise Grammatik, Rechtschreibung, die Textlänge oder die Wortanzahl des Schreibprodukts. Rückmeldungen auf der hierarchieniedrigeren Textebene sind leicht zu beheben und daher einfach zu implementieren. Gleichwohl haben diese einen nicht so starken Einfluss auf die Textqualität, da sie eher Oberflächenüberarbeitungen adressieren und daher weniger substantielle Aspekte von Textqualität tangieren (Patchan et al., 2016; Strobl et al., 2019). Im Gegensatz dazu wird angenommen, dass Feedback auf hierarchiehöherer Textebene einen größeren Einfluss auf die Textqualität hat (Patchan et al., 2016). Rückmeldungen auf hierarchiehöherer Textebene beziehen sich beispielsweise auf die Struktur und Organisation, den Stil, die Gliederung oder die Kohäsion des Schreibprodukts. Durch die Information über substantielle Aspekte eines aktuellen Schreibprodukts können Schreibende angeregt werden, solche Revisionen zu realisieren, die die Textqualität nachhaltig verbessern. Die Rückmeldung auf hierarchiehöherer Textebene ist insbesondere sinnvoll, da Lernende dazu

tendieren, oberflächliche Korrekturen und Überarbeitungen auf hierarchieniedrigerer Ebene vorzunehmen (Hayes et al., 1987; Lachner et al., 2017a; Patchan et al., 2016) und substanzielle Überarbeitungen auszusparen.

2.5.2 Dimension der Repräsentation

Neben der Dimension der Textqualität lassen sich verschiedene Repräsentationsformen unterscheiden. Als relativ einfaches Verfahren haben sich *numerische* Repräsentationen herausgestellt, welche Feedbackinformation direkt in Form von Zahlen (Score, Note, erreichte Punktezahl) wiedergeben (Kellogg et al., 2010; Roscoe, Brandon et al., 2013; Wilson & Roscoe, 2020). Oft werden diese auch in Form von Diagrammen, Symbolen und Ratingskalen wiedergegeben. Beide Verfahren eignen sich tendenziell eher als summative Rückmeldung.

Neben diesen einfacheren Repräsentationsformaten gibt es zudem *eingebettete Formate*, welche Feedbackinformation direkt innerhalb des Textes, bspw. durch Hervorhebungen (Burkhart et al., 2020; Moore & MacArthur, 2016; Tsai et al., 2020), darstellen und dadurch eine leichtere Integration des Feedbacks in die Schreibrepräsentation ermöglichen. Diese eingebetteten Formate eignen sich sehr gut für die formative Rückmeldung auf Textentwürfe. Neuere Ansätze erlauben zusätzlich *piktorale Repräsentationen* eines Textes, in denen die Feedbackinformation bspw. als Graph repräsentiert wird, z. B. in Form einer Concept Map (Kim, 2012; Lachner et al., 2017a, 2017b; O'Rourke et al., 2011; Pirnay-Dummer & Ifenthaler, 2011; Villalon & Calvo, 2011).

2.5.3 Dimension der Spezifität

Bei der Art und Weise, wie ausführlich bzw. umfassend Feedback gegeben wird, kann zwischen einer holistischen und einer analytischen Bewertung unterschieden werden. Im Rahmen einer holistischen Rückmeldung wird das Schreibprodukt als Ganzes betrachtet. Die Bewertung erfolgt anhand eines Gesamtindikators, der sich unter Umständen aus

mehreren Teilaspekten von Textqualität zusammensetzt (vgl. Schipolowski & Böhme, 2016). Wie aktuelle Forschung zeigt, ist insbesondere die Verwendung holistischer Scores unter den computerbasierten Rückmeldesystemen verbreitet (Strobl et al., 2019). Allerdings ist holistisches Feedback durch seine große Allgemeinheit oft für Lernende nur schwer fassbar, was ggf. in eine geringere Implementation des Feedbacks in der Überarbeitung mündet (siehe MacArthur, 2015; Roscoe & McNamara, 2013; Sturm, 2014, 2016). Die Ergebnisse der Metaanalyse von Wisniewski et al. (2020) bestätigen abermals, dass Feedback effektiver ist, je mehr Informationen es enthält und dass allgemeines, korrigierendes Feedback mit wenig Informationsgehalt kleine Effekte aufweist, wohingegen ausführlicheres bzw. spezifischeres Feedback (auf Aufgaben-, Prozess- und Selbstregulationsebene) am effektivsten ist (Hattie & Timperley, 2007; Wisniewski et al., 2020). Lachner et al. (2017a) verglichen in einer experimentellen Feldstudie unterschiedliche Arten holistischen und spezifischen Feedbacks bezüglich der Textkohärenz der studentischen Schreibentwürfe im Rahmen einer Vorlesung ($N = 251$). Die Autoren konnten ebenfalls zeigen, dass es Schreibenden subjektiv schwerer fiel, holistisches Feedback in ihrer Überarbeitung zu implementieren. Der höhere Grad an subjektiv wahrgenommener Schwierigkeit mündete insgesamt in eine niedrigere Schreibperformanz (gemessen an dem Grad der Kohäsion). Die Befunde verdeutlichen, dass es zusätzlicher Unterstützung bedarf, um holistische Feedbackverfahren zu verarbeiten und umzusetzen. Hier könnten Leitfragen helfen, die Effektivität von holistischem Feedback zu erhöhen (siehe McNamara et al., 2015, oder Roscoe & McNamara, 2013, für empirische Umsetzungen anhand des Feedbacksystems Writing Pal). Im Gegensatz zu holistischer Rückmeldung wird bei einem analytischen Beurteilungsverfahren spezifische Rückmeldung auf verschiedene Kriterien des Schreibprodukts gegeben (Schipolowski & Böhme, 2016). Durch solch ein kriteriengeleitetes Vorgehen ist eine differenziertere Betrachtung des Schreibprodukts möglich und Förderbedarfe bezüglich der verschiedenen

Anforderungsbereiche der Schreibkompetenz werden ersichtlich. Durch diese Transparenz können Lernende die Beurteilung ihres Textes nachvollziehen und die noch ausstehenden Lernziele werden für die Lernenden konkreter fassbar. Dies ist vor allem für die Überarbeitung des Textentwurfs hilfreich. Ähnliche Befunde ließen sich zum Beispiel auch bei Peer-Feedback feststellen. Patchan et al. (2016) analysierten in ihrer Studie mehr als 7'500 Rückmeldungen von 351 Peers und setzten diese in Bezug zur Implementation und Qualität der Überarbeitung. Die Autor*innen konnten zeigen, dass Rückmeldungen leichter implementiert werden konnten, wenn diese sich auf differenzierte Schwächen im Text bezogen (vgl. auch Hattie & Timperley, 2007; Wisniewski et al., 2020). Die Qualität der Revision konnte insbesondere durch den Fokus auf substantielle Kommentare gesteigert werden. Die Befunde verdeutlichen, dass holistische Rückmeldung eher schwer von Lernenden verarbeitet werden kann. Analytische Methoden dagegen scheinen dagegen für formatives Feedback geeignet zu sein, da durch die Spezifität distinkte Revisionen beim Schreiben angeregt werden können. Die Spezifität kann hierbei noch optimiert werden, beispielsweise durch die Zugabe von Hinweisreizen, die potenzielle Defizite lokalisieren und Vorschläge zur Verbesserung geben (Patchan et al., 2016).

2.6 Evidenzbasierte Feedbacksysteme zur Förderung von Schreibkompetenzen

Im Folgenden werden drei in den letzten Jahren entwickelte Feedbacksysteme vorgestellt (Criterion, Writing Pal, CohViz) und hinsichtlich der oben genannten Dimensionen eingeordnet. Diese Feedbacksysteme wurden ausgewählt, da sie sich durch eine systematische Begleitforschung auszeichnen und entsprechend Aussagen über die Effektivität erlauben.

Table 1*Computerbasierte Feedbacksysteme im Vergleich hinsichtlich der postulierten**Hauptdimensionen*

Tool	Textebene	Repräsentationsform	Spezifität	Sonstiges
Criterion	Hauptsächlich auf hierarchie-niedriger Ebene	Holistischer Score auf einer 6-Punkte-Skala, der sich aus 6 Werten zusammensetzt	<ul style="list-style-type: none"> – Keine Lösungshinweise – Nicht lokalisiert 	Lehrpersonen und Peers können auch Feedback geben und Kommentare hinterlassen.
Writing Pal	Hauptsächlich auf hierarchie-höherer Ebene	Holistischer Score auf einer 6-Punkte-Skala + Leitfragen und Prompts	<ul style="list-style-type: none"> – Lösungshinweise vorhanden – nicht lokalisiert 	Lehrpersonen können zusätzliche Informationen über das System bereitstellen.
CohViz	Ausschließlich auf hierarchie-höherer Ebene (Kohäsion)	Concept Map + Leitfragen	<ul style="list-style-type: none"> – Leitfragen (Lösungshinweise) vorhanden – lokalisiert 	Keine Interaktionsmöglichkeit für Lehrpersonen oder Peers.

2.6.1 Criterion

Criterion ist im anglo-amerikanischen Raum ein viel eingesetztes und evidenzbasiertes Feedbackinstrument. Bei Criterion handelt es sich um ein webbasiertes e-rater-System, welches von ETS, der größten privaten Organisation für Bildungstest und -messungen, entwickelt wurde. E-rater-Systeme liefern diagnostisches Feedback anhand der e-rater scoring engine, einer Anwendung des NLP, bei der englische Texte der Lernenden anhand von Merkmalen mit Musteraufsätzen, die von geschulten Leser*innen und Dozent*innen bewertet wurden, verglichen werden. Auf dieser Grundlage gibt das System einen holistischen Score aus, der den Lernenden als Leitfaden für die Überarbeitung ihres Textes dienen soll.

Der holistische Score, der die Gesamtqualität des Textes widerspiegelt, setzt sich zusammen aus sechs Werten auf jeweils einer Skala von eins bis vier oder eins bis sechs,

wobei die Eins für den niedrigsten Wert steht. Die sechs Werte beziehen sich auf die Bereiche Grammatik, Rechtschreibung, Mechanik, Gebrauch, Organisation und Entwicklung des Textes. Da es sich bei Criterion um einen Onlineservice handelt, kann dieses Feedbacksystem unter der Voraussetzung eines Internetanschlusses von überall genutzt werden. Allerdings ist Criterion lizenziert und nur im Rahmen von formellen Lernkontexten nutzbar, wenn Lehrpersonen über eine Lizenz verfügen. Criterion kann in allen Phasen des Schreibens sofortiges, automatisches Feedback bereitstellen, es bewertet aber nicht den Inhalt, die Argumentationsstruktur oder die Kohäsion des zu bewertenden Textes. Criterion ist demnach eher als Ergänzung zur Anleitung und Rückmeldung von Lehrpersonen geeignet und dient auch dazu, Lehrende zu entlasten und Unterrichtszeit zu sparen, sodass die Lehrpersonen die Lernenden individuell beraten und unterstützen können. Trotzdem setzt die lernwirksame Verwendung von Criterion eher motivierte Lernende und gut informierte Lehrende voraus. Lehrpersonen können durch das Tool aber auch anders unterstützt werden, indem sie Lernenden über den Service von Criterion Schreibstrategien und Unterrichtsinhalte vermitteln, deren Aufsätze bewerten und mit ihnen direkt in Kontakt treten können. Dabei können eigene Schreibaufgaben und Themen erstellt oder die Criterion-Themenbibliothek verwendet werden. Es können aber nicht nur die Lehrpersonen, sondern auch Peers (einzeln oder in Gruppen) oder das System selbst automatisiert Feedback auf die von den Lernenden erstellten Textentwürfe geben.

Im Vergleich zu menschlichen Bewerter*innen zeigte das Criterion-Bewertungssystem sehr hohe Übereinstimmungen (Chodorow et al., 2010). Darüber hinaus konnte in mehreren Studien die Wirksamkeit von Criterion dokumentiert werden. Lernende, die ihren Text mit Criterion-Feedback überarbeitet haben, machten im Vergleich zur Kontrollgruppe ohne Feedback signifikant weniger Fehler in ihren überarbeiteten Aufsätzen und konnten dadurch ihre Schreibergebnisse verbessern (Attali, 2004; Attali & Burstein, 2006; Chodorow et al., 2010).

Beispielsweise untersuchten Kellogg et al. (2010), ob sich die Schreibleistung von Studierenden durch die Anwendung von Criterion-Feedback über drei Wochen verbesserte. Im Rahmen der Studie wurden Studierende zufällig in eine von drei Gruppen eingeteilt und schrieben in drei aufeinander folgenden Wochen jeweils einen Essay. Die erste Gruppe erhielt lediglich zum zweiten Essay Rückmeldung, die zweite Gruppe bekam Feedback auf alle drei Essays und eine dritte Gruppe erhielt als Wartekontrollgruppe Feedback erst nach Beendigung der Studie. Nach einem Verzögerungsintervall von zwei Wochen wurde ein Testtext ohne Feedback geschrieben, um den Lerntransfer zu beurteilen. Im Ergebnis zeigte sich, dass Studierende aus der zweiten Gruppe, die Feedback zu jedem Essay erhalten hatten, bessere Schreibprodukte produzierten als die Studienteilnehmer*innen aus den beiden anderen Gruppen. Die Befunde illustrieren, dass die Kontinuität des Feedbacks bedeutsam ist, um die Effektivität des Feedbacks zu verbessern.

2.6.2 Writing Pal

Writing Pal ist ein webbasiertes Software-Tool aus dem Bereich der intelligenten Tutorensysteme (ITS) und wurde vom Science of Learning and Educational Technology (SoLET) Lab an der Arizona State Universität auf Basis nutzen-inspirierter Grundlagenforschung entwickelt (Crossley et al., 2013; Roscoe, Brandon et al., 2013; Roscoe & McNamara, 2013; Roscoe et al., 2015). Ziel von Writing Pal ist es, Lernenden Strategieinstruktionen (z. B. anhand von Unterrichtsvideos) in Form eines Gamificationansatzes (siehe Bai et al., 2020, für einen Überblick) zu vermitteln und ihnen die Möglichkeit zu bieten, englischsprachige Aufsätze automatisiert auf bestimmte Textmerkmale hin (wie z. B. Kohäsion, Länge des Textes, Ausarbeitung oder Relevanz des Themas) zu überprüfen. Im Hinblick auf die Ebene der Textqualität stellt Writing Pal Feedback sowohl auf hierarchieniedriger (z. B. Textlänge, Wortanzahl) als auch auf hierarchiehöherer Ebene (z. B. Kohäsion) bereit. Die ganzheitliche Qualität der Aufsätze

wird durch Algorithmen zur Verarbeitung natürlicher Sprache (NLP) bewertet, die lexikalische, syntaktische, semantische und rhetorische Merkmale des Schreibprodukts berücksichtigen (McNamara et al., 2013, 2015). Diese Algorithmen generieren dann einen holistischen Gesamtscore der Textqualität (6-Punkte-Skala). Um Revisionen anzuregen, bietet Writing Pal zusätzlich adaptive Leitfragen, basierend auf dem Gesamtscore, und Methoden zur Verbesserung des Textes (Roscoe & McNamara, 2013). Bei diesem formativen Feedback wird der Fokus auf die hierarchie-höhere Textebene gelegt. Die Lernenden erhalten bei jedem Textentwurf zunächst Rückmeldung in Form einer Leitfrage oder einer prompt-ähnlichen Empfehlung zu einem diagnostizierten Schreibproblem ihres Schreibprodukts. Anschließend können die Lernenden ihren Textentwurf daraufhin überarbeiten und erneut Feedback zu dieser Textstelle bzw. diesem Schreibproblem erhalten oder Feedback zu einem weiteren Thema anfordern. Neben Aufgaben, die durch das System vorgeschlagen werden, haben Lehrpersonen ebenfalls die Möglichkeit, den Lernenden über das System selbst Informationen bereitzustellen oder die Lernumgebung anzupassen. In erster Linie ist das System nicht als Ersatz für den Unterricht oder zum Erledigen der Hausaufgaben gedacht. Stattdessen soll es durch geführte Anleitungen ergänzend zum Unterricht als Schreibübung zur Verbesserung der Schreibfähigkeiten dienen. Trotzdem können Lernende Writing Pal auch explizit nutzen, um ihre Schreibfähigkeiten in bestimmten Bereichen der schriftlichen Textproduktion, in denen sie noch Schwierigkeiten haben, anhand gezielter Aufgaben individuell zu üben.

Writing Pal hat sich als sehr valides System erwiesen, da die systemgenerierten Feedbackscores eine hohe Übereinstimmung mit menschlichen Expert*innenratings aufwiesen (McNamara et al., 2015). Darüber hinaus konnte in zahlreichen (Längsschnitt-)Studien (z. B. McNamara et al., 2015; Roscoe & McNamara, 2013; Roscoe et al., 2015) gezeigt werden, dass Lernende mit Writing Pal ihre Schreibfähigkeiten (im Laufe der Zeit) verbessern. Zum Beispiel untersuchten Roscoe und McNamara (2013) den Einfluss von

Writing Pal unter anderem auf die Essayqualität von 141 US-amerikanischen Zehntklässler*innen in einem Unterrichtszeitraum von sechs Monaten. Alle Schüler*innen schrieben auf Basis vorgegebener Prompts Essays und hatten dann die Möglichkeit diese mithilfe des Writing Pal-Feedbacks zu überarbeiten. Die Schüler*innen konnten ihre Schreibleistung während der Nutzung signifikant verbessern. Allerdings ist offen, da das Design keine Kontrollgruppe berücksichtigte, ob diese Zuwächse aufgrund einfacher Zeiteffekte oder aus der Nutzung von Writing Pal resultierten. Zudem monierten die teilnehmenden Lernenden, dass sie Schwierigkeiten mit der Allgemeingültigkeit bzw. der Genauigkeit der im Feedback bereitgestellten Informationen hatten und es deshalb schwer umzusetzen war (Roscoe & McNamara, 2013).

2.6.3 CohViz

CohViz (<https://github.com/ch-bu/cohviz>) basiert ebenfalls auf aktuellen computerlinguistischen Methoden und Visualisierungstechniken und nutzt NLP-Technologien. Im Gegensatz zu den beiden oben genannten Systemen verfolgt CohViz jedoch einen piktoralen Ansatz, indem es automatisch grafische Repräsentationen von Texten in Form einer Concept Map generiert (siehe Figure 5). CohViz eignet sich vorrangig für tendenziell eher kurze Texte (< 300 Wörter), wie z. B. Erklärungen oder Erörterungen. Allerdings ist CohViz eines der wenigen Tools, das Rückmeldung sowohl auf deutsche als auch auf englische Texte gibt. Hinsichtlich der Dimension der Textqualität stellt CohViz ausschließlich Feedback auf die Textkohäsion (hierarchie-höhere Ebene) bereit. Kohäsion wird in der Fachwelt unterschiedlich operationalisiert. Generell sind Kohäsionsmittel linguistische Mittel, um Textteile und Sinnzusammenhänge eines Textes zu verdeutlichen. In diesem Beitrag definieren wir Kohäsion aus einer eher psychologischen Perspektive und unterscheiden zwischen lokaler und globaler Kohäsion. Lokale Kohäsion bezieht sich auf die Satzebene eines Textes. Lokale Kohäsionsmittel werden zum Beispiel genutzt, um

Evidenzbasierte Feedbacksysteme zur Förderung von Schreibkompetenzen

anhand von syntaktischen Verknüpfungen (z. B. Verwendung von Konnektiven) oder durch semantische Verbindungen (z. B. Nutzung von Wortwiederholungen) inhaltliche Sinnzusammenhänge zwischen benachbarten Sätzen herzustellen (McNamara et al., 2013). Globale Kohäsion bezieht sich hingegen auf den übergeordneten Sinnzusammenhang zwischen einzelnen Textteilen und dem Gesamttext. Zu den globalen Kohäsionsmitteln gehört beispielsweise die Einteilung des Textes in Absätze und der Einsatz von Überschriften, sowie die Berücksichtigung der rhetorischen und logisch-semantischen Struktur des jeweiligen Textgenres (z. B. Einführung zentraler Konzepte zu Beginn einer Erklärung; Lachner et al., 2017b). CohViz gibt sowohl Feedback im Hinblick auf die lokale als auch auf die globale Kohäsion. Die Knoten in der Concept Map visualisieren die einzelnen Konzepte eines Textes (globale Kohäsion); die Pfeile bezeichnen die semantischen Verbindungen zwischen den Konzepten (lokale Kohäsion).

In einer breit angelegten Validierungsstudie mit einem Textkorpus von über 1'000 Texten aus der deutschen Wikipedia-Datenbank konnten hohe Übereinstimmungen sowohl mit menschlichen Ratern als auch etablierten Maßen (z. B. Argumentüberlappung) gezeigt werden (Burkhart et al., 2020). Im Gegensatz zu Criterion und Writing-Pal gibt CohViz keine direktive Rückmeldung in Form von Verbesserungsvorschlägen, sondern soll Studierende durch die Visualisierung der Diskursstruktur und der inhaltlichen Bedeutung der eigenen Erklärung für distinkte strukturelle und semantische Merkmale (bspw. lokale Kohäsion, globale Kohäsion, inhaltliche Adäquatheit) sensibilisieren und anregen, über bestimmte inhaltliche und diskursive Probleme ihrer Schreibprodukte nachzudenken. Diese Reflexionsprozesse werden mit Hilfe didaktischer Leitfragen (Prompts) verstärkt, welche kurze Aufforderungen enthalten, wie die Concept Map zur Verbesserung der eigenen Erklärung genutzt werden sollte (siehe Figure 5).

Figure 5

Beispiel einer in CohViz generierten Concept Map zum Thema Osmose

Für deine Überarbeitung erhältst du im Folgenden eine Concept Map deines Textes. Diese Concept Map stellt eine grafische Visualisierung deiner Erklärung dar: Die zentralen Konzepte deiner Erklärung werden als Kreis dargestellt und deren Verbindungen als Pfeile. Konzepte, die keine Verbindungen zu den restlichen Aussagen haben stellen lokale Kohäsionsbrüche dar.

Überprüfe die globale Kohäsion deiner Erklärung:

1. Wurde ausreichend für die Lesenden in das Thema eingeführt?
2. Sind alle wichtigen Konzepte und Mechanismen des Phänomens für die Lesenden erklärt?
3. Wurde das Phänomen hinreichend für die Lesenden durch Beispiele illustriert?

Überprüfe die lokale Kohäsion deiner Erklärung:

1. Wie können die dargestellten lokale Kohäsionsbrüche geschlossen werden, um die Verständlichkeit für die Lesenden zu erhöhen?
2. Sind zwischen den Sätzen ausreichend Überleitungen für die Lesenden gegeben?
3. Kann von den Lesenden leicht eine Verbindung zwischen den Sätzen der Erklärung hergestellt werden?

Als Osmose wird in den Naturwissenschaften der gerichtete Fluss von molekularen Teilchen durch eine selektiv- oder semipermeable Trennschicht bezeichnet.

Das Phänomen wird im Rahmen der statistischen Mechanik theoretisch erklärt. Die physikalischen Abläufe auf mikroskopisch-molekularer Ebene sind zu Beginn des 21. Jahrhunderts Gegenstand von Gelehrtenstreit und aktiver Forschungstätigkeit.

Anmerkung. Oben befindet sich die Instruktion inklusive Prompts. Links unten ist der Erklärtextentwurf. Rechts unten befindet sich die dazugehörige Concept Map.

Kohäsionsbrüche werden sichtbar durch unverbundene Fragmente der Concept Map, die unterschiedlich farblich dargestellt werden.

CohViz wurde bisher anhand mehrerer experimenteller Studien validiert. Die Effektivität konnte in mehreren Studien dokumentiert werden, da Studierende, die CohViz-Feedback erhielten, kohäsivere und insgesamt verständlichere Erklärungen generierten (Burkhart et al., 2020; Lachner et al., 2017a, 2017b; Lachner & Neuburg, 2019). Eine Analyse der zugrundeliegenden Prozesse mittels Laut-Denken-Protokollen und Eyetracking zeigte, dass die Analyse der Concept Maps bei Studierenden insbesondere Integrationsprozesse zwischen dem eigenen Text und der Concept Map förderte und sie dahingehend anregte, über die globale Struktur ihrer Erklärung nachzudenken und geeignete lokale und globale

Kohäsionsstrategien zu planen, um diese Kohäsionsdefizite zu schließen. Weitere Forschung ist jedoch noch nötig, um solche Tools für längere Textsorten zu adaptieren.

Im Gegensatz zu CohVz, das piktorales Feedback für englische und deutsche Texte generiert, sind Criterion und Writing Pal beide für den englischsprachigen Raum entwickelt worden und stellen Feedback in Form eines holistischen Scores bereit. Aktuelle Studien zeigen jedoch, diese Art der Feedbackrepräsentation nur wenig verwertbare Informationen für die Lernenden zur Verfügung stellt und das Feedback deshalb schwerer umsetzbar ist (MacArthur, 2015; Roscoe & McNamara, 2013; Sturm, 2016; Wisniewski et al., 2020). Deshalb sollten zusätzlich zu einem holistischen Score weitere Informationen zur Revision der Textentwürfe bereitgestellt werden. Außerdem könnten zusätzliche Leitfragen und Lösungshinweise (wie z. B. bei Writing Pal oder CohViz) und die Lokalisierung der Rückmeldungen (in CohViz möglich) die Umsetzung des Feedbacks für die Lernenden unterstützen. Des Weiteren ist beim Einsatz der Tools zu beachten, welche Voraussetzungen die Lernenden mitbringen müssen, um effektiv damit arbeiten zu können (z. B. selbstregulierte Interpretation des Feedbacks bei CohViz). Grundsätzlich erwiesen Criterion, Writing Pal und CohViz hohe Übereinstimmungen mit menschlichen Rater*innen (Burkhart et al., 2020; Chodorow et al., 2010; McNamara et al., 2015), was für eine hohe Validität der Technologien spricht. Gleichzeitig zeigten alle Technologien eine relative Wirksamkeit, um Schreibstrategien zu fördern (z. B. Attali & Burstein, 2006; Burkhart et al., 2020; Chodorow et al., 2010; Lachner et al., 2017a, 2017b; Lachner & Neuburg, 2019; McNamara et al., 2015; Roscoe, Brandon et al., 2013). Insgesamt ist beim Einsatz computerbasierter Technologien zu beachten, dass diese nicht als Ersatz für den Unterricht, sondern als Ergänzung zu Unterricht oder Rückmeldungen von Lehrpersonen gesehen werden sollten. Eine weitere Empfehlung, die die Studien zu den genannten Tools gemein haben ist, dass ein kontinuierlicher Einsatz der Tools empfohlen wird. Denn oft zeigen sich insbesondere über einige Wochen der Toolnutzung hinweg positive Effekte auf die Textqualität und die

Verbesserung der Schreibkompetenz der Lernenden (Kellogg et al., 2010; Roscoe & McNamara, 2013).

Wenn Lehrpersonen digitale Medien in Form von computerbasierten Feedbacktools in ihrem Schreibunterricht einsetzen wollen, sollten sie sich an folgenden Leitfragen orientieren (basierend auf den Feedbackdimensionen aus Kapitel 2.5):

- Welchen Zweck soll das Tool erfüllen?
- Auf welche Textebene soll sich das Feedback beziehen (niedrige oder hohe Ebene oder beides)?
- Wie spezifisch soll die Rückmeldung für die Lernenden sein? Davon ist meist auch die Repräsentationsform abhängig. Also, wie soll das Feedback dargestellt werden (nur holistischer Score, piktoral oder eine Kombination verschiedener Repräsentationsformen)?
- In Abhängigkeit von der Repräsentationsform stellt sich die Frage, ob zusätzliche Kompetenzen (z. B. Selbstregulation) von den Lernenden vorausgesetzt werden?
- Soll das Tool im Rahmen des Unterrichts interaktiv (z. B. in Form von Kommentaren der Peers und der Lehrperson) als Ergänzung zu den Anleitungen und zur Entlastung der Lehrperson verwendet werden (wie z. B. bei Criterion und Writing Pal)?
- Oder soll das Tool eine Möglichkeit bieten, Schreibstrategien nicht nur zu erlernen und zu üben, sondern auch zu vermitteln (wie z. B. bei Writing Pal anhand von Strategieinstruktionen in Form von Unterrichtsvideos und Gamificationelementen)?
- Je nach Lizenz und Verfügbarkeit kann das Tool auch eher zu Übungszwecken verwendet werden (wie z. B. CohViz oder Writing Pal).

Wie sich exemplarisch an den drei vorgestellten Tools Criterion, Writing Pal und CohViz zeigt, existieren bereits sehr wirksame Tools, um Lehrpersonen in ihrem Schreibunterricht

Forschungsdesiderate und weitere Entwicklung

zu unterstützen und Lernende bei der Verbesserung ihrer Schreibkompetenzen und der Qualität ihrer Textentwürfe zu fördern. Nichtsdestotrotz zeigt die Beantwortung der Leitfragen und die Beschreibung der Tools, dass es bisher auch kein Tool gibt, das alle Anforderungen gleichermaßen erfüllt. Demnach ist die weitere Entwicklung und Erforschung von (neuen) computerbasierten Feedbacktools weiterhin von Nöten.

2.7 Forschungsdesiderate und weitere Entwicklung

In diesem Beitrag wurde der Versuch unternommen, das Potenzial digitaler Medien mit Fokus auf computerbasiertem Feedback zu konzeptualisieren. Hierzu wurden drei verschiedene forschungsbasierte Feedbacksysteme basierend auf den Dimensionen der Textqualität, der Repräsentationsform und der Spezifität des Feedbacks vorgestellt. Die exemplarischen Feedbacksysteme illustrieren, dass computerbasiertes Feedback durchaus eine effektive und evidenzbasierte Methode sein kann, um Lernende in den verschiedenen Phasen des Schreibprozesses (Planung, Entwurf und Revision) zu unterstützen. Aus technologischer Perspektive sind jedoch die Entwicklung und Erforschung weiterer Tools nötig, um insbesondere Feedback auf hierarchiehöherer Ebene bereitzustellen. Hierzu könnten neuere Ansätze der künstlichen Intelligenz (z. B. in den Bereichen Learning Analytics, Machine Learning oder Educational Data Mining) helfen, um solche Feedbacksysteme zu optimieren. Darüber hinaus wäre die Entwicklung von weiteren Tools wünschenswert, die auch andere Sprachen außer Englisch unterstützen. Neben diesen technologischen Entwicklungen ist insbesondere weitere Forschung zu den Implementationskontexten erforderlich. Bisher existieren hauptsächlich Machbarkeits-, Effektivitäts- sowie Validierungsstudien. Auf dieser Grundlage lassen sich jedoch so gut wie keine direkten Schlüsse auf die Anwendung in der Unterrichtspraxis und die langfristige Effektivität der Systeme ziehen. Eine stärkere Integration in den Unterricht und eine Orchestrierung von Feedbacksystemen mit anderen Lehr- und Lernmethoden ist daher

wünschenswert (Strobl et al., 2019). Aus methodischer Perspektive wären zudem randomisierte, kontrollierte Feldstudien (randomized controlled field trials) ein fruchtbarer Ansatz, um empirische Nachweise über die Effektivität von Feedbacksystemen im Feld zu generieren (siehe Meurers et al., 2019 für den Zweitspracherwerb). Idealerweise sollten diese insbesondere Verläufe des Feedbackesinsatzes dokumentieren, da der Einsatz digitaler Medien oft sogenannten Neuigkeitseffekten unterliegt (Clark, 1983) oder Wirksamkeit erst durch mehrmalige Nutzung resultiert (Kellogg et al., 2010). Eine Kombination beider Ansätze kann somit zur Entwicklung evidenzbasierter Feedbacksysteme beitragen und angehenden Schreibenden helfen, genuine Schreibkompetenzen zu erwerben

3 DOES COMPUTER-BASED FEEDBACK FOSTER EXPOSITORY WRITING?

Wagner, S., Schneider, J., Backfisch, I., Scheiter, K., & Lachner, A. (in revision). Does Computer-Based Feedback Foster Expository Writing? A Meta-Analysis.

This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. The Version of Record of this manuscript is currently invited for major revision in the journal *Educational Psychology Review*.

Abstract

Feedback is a beneficial instructional strategy to improve students' writing. To this end, computer-based feedback technologies are discussed as promising supplements in addition to instructor feedback. Several computer-based feedback approaches have been developed. However, to date, if and under which conditions computer-based feedback can be effective is unclear. We conducted a meta-analysis to investigate the effectiveness of computer-based feedback on writing quality, and to explore its boundary conditions ($k = 24$ studies with pre-posttest design; $k = 15$ studies with pretest-posttest-control group design). The results suggest that computer-based feedback contributed to writing quality with a small to medium effect ($g = .53$), but only for studies with pretest-posttest design. For studies with pretest-posttest-control group design, the effect ($g = .27$) was not significant. Moderator analyses revealed that firstly text level was a significant moderator: Computer-based feedback was only effective for enhancing surface features of the text (e.g., grammar, spelling), as such feedback was likely easier to implement. Secondly, feedback specificity was a strong moderator as specific computer-based feedback improved students' text quality more than generic feedback. The results highlight the need for the development of sophisticated technologies to provide students with high-level feedback and subsequently scaffold their expository writing.

Keywords: computer-based feedback, writing quality, meta-analysis, technology-enhanced learning and instruction

3.1 Introduction

Writing comprehensible and elaborated expository texts is regarded as a key 21st-century skill (Concha & Paratore, 2010; Geiser & Studley, 2001; Lachner et al., 2017a; Powell, 2009; Sharp, 2016). When producing comprehensible texts, students need to iteratively plan, execute, and revise their writing (e.g., Hayes & Flower, 1986; Kellogg & Whiteford, 2009), which puts high demands on students. Feedback has been demonstrated as a beneficial instructional strategy to particularly support students' writing processes, as it enhances processes of monitoring and regulation (Graham et al., 2015; Kellogg & Raulerson, 2007; Kieft et al., 2008; Philipp, 2017; Roscoe & McNamara, 2013). Although feedback has been demonstrated to be effective, it is regarded to be rather time-consuming and laborious (Brindle et al., 2016; Graham, 2019; Graham & Herbert, 2011). To this end, computer-based feedback has been discussed as a potential supplement to support students' writing (Kellogg & Whiteford, 2009; Roscoe & McNamara, 2013; Strobl et al., 2019). To date, several feedback technologies have been developed. However, meta-analytic evidence is missing, that allows to draw conclusions regarding whether and under what conditions computer-based feedback is generally effective for enhancing students' writing.

To close this research gap, we conducted a meta-analysis to 1) estimate the overall effectiveness of computer-based feedback in the context of expository writing, 2) as well as examining potential boundary conditions. Therefore, we particularly focused on boundary conditions which may have resulted from system-related design aspects of the computer-based feedback (i.e., the specificity, the format of representation, or the type of text level that was addressed within the feedback), as well as individual boundary conditions, based on students' prerequisites (i.e., prior knowledge).

3.2 Expository Writing

Writing can be regarded as a self-regulated process (e.g., Harris et al., 2008; Hayes, 2012; Hayes & Flower, 1986; Philipp, 2014). Hayes (2012) proposed three different self-regulation levels which constitute effective writing: (1) The *control level* comprises a writer's motivation including the goals and plans of the subsequent writing process and the product, as well as her or his rhetorical and content-related prior knowledge (Becker-Mrotzek et al., 2014; Bereiter & Scardamalia, 1987). (2) The *process level* comprises different sub-processes to achieve the writing goals: First, a writer proposes a writing plan, translates the proposed ideas into propositions and subsequently transcribes them into a written draft (Hayes & Olinghouse, 2015). Based on the draft the writer ideally iteratively monitors and revises the draft to create a final writing product which ideally attains the previously set writing goals. (3) The *resource level* consists of generic resources, such as attention, working memory resources, and pre-requisite skills, such as reading.

The model by Hayes (2012) illustrates that writing is a complex task, requiring assistance during early stages of writing (Becker-Mrotzek et al., 2014; Bereiter & Scardamalia, 1987; Flower & Hayes, 1980; Hayes & Flower, 1986; Kellogg & Whiteford, 2009). The need for additional support, particularly holds true for revision processes, as seminal studies indicated that – compared to experienced writers – novice writers less frequently engage in revision activities that tackle substantive aspects of their writing (see Chanquoy, 2009; Cho & MacArthur, 2010; Lachner et al., 2017b; MacArthur & Graham, 2016; Roscoe et al., 2016).

3.3 Feedback to Support Revision Processes

Formative feedback is one of the most effective strategies for stimulating students' revision activities and enhance writing quality (Graham et al., 2015; Philipp, 2017). Commonly, formative feedback is understood as information about distinct aspects of a learner's current performance (e.g., writing quality). The goal of feedback is to provide

information to help the recipient reduce discrepancies between the actual and the targeted performance (Hattie & Timperley, 2007; Narciss, 2008; Stevenson & Phakiti, 2014; Winne & Butler, 1994). Formative feedback, thus, not only includes assessment aspects, but also information on how to regulate the further learning process and improve their current performance (i.e., quality of the writing product) regarding assessment criteria (Butler & Winne, 1995; Narciss, 2008).

Recent meta-analyses showed that formative feedback can support students in developing their writing skills. For example, Graham et al. (2015) examined effects of generic formative assessment and feedback on children's (kindergarten to grade 8) writing quality based on 16 empirical studies. The authors obtained a medium to large effect ($d = 0.77$) of formative feedback on writing quality compared to another treatment condition or a control group that did not receive feedback (see also Wisniewski et al., 2020, for related findings on the general effectiveness of feedback).

3.4 Computer-Based Feedback – A Feasible Approach to Foster Expository Writing?

Given that formative feedback is often regarded as time-consuming and laborious, several computer-based feedback technologies have been proposed as a supplement to instructor feedback (e.g., Allen et al., 2016; Kellogg & Whiteford, 2009; Roscoe & McNamara, 2013). For instance, automated writing evaluation systems (AWE), such as Criterion, use natural language processing (NLP) technologies, to provide feedback to students and initiate distinct revision activities (Dikli, 2006; Shermis & Burstein, 2003; Wilson & Roscoe, 2020; see also Lachner et al., 2017a; Roscoe & McNamara, 2013, for other examples of tools using NLP). AWE-systems, therefore, synthesize feedback on several dimensions, ranging from low-level features, such as grammar or spelling to high-level features, such as the cohesion or the complexity of a text. In most cases, the AWE system generates a holistic score across these dimensions, which is then automatically provided as feedback to the learner. This

feedback is represented in different formats, such as numeric, text-based, or graph-based information.

Similarly, other feedback systems, for instance, Intelligent Tutoring Systems (ITS; e.g., Summary Street, Writing Pal) use algorithms which compare students' drafts with writing samples which were assessed by writing experts. ITS also combine different representation formats of feedback, e.g., written comments/text-based feedback and additional scores when providing feedback to learners. Yet, other feedback tools (e.g., CohViz) use graphical approaches, based on network analysis and decision trees to provide graphical feedback regarding the cohesion of students' texts (Burkhart et al., 2020).

Strobl et al. (2019) conducted a systematic review to map the research landscape on computer-based feedback tools for writing support. The authors showed that there is a wide variety of existing tools that differ in terms of system-related aspects of the feedback (e.g., specificity of the provided information or the targeted text level), but also regarding the learner-related settings (e.g., knowledge level, educational level), in which the feedback was implemented. Most of the other previous research syntheses investigating computer-based feedback on writing focused exclusively on AWE systems. For instance, Nunes et al. (2021) and Fleckenstein, Reble et al. (2023) conducted systematic reviews on AWE feedback and identified potential influencing factors (e.g., role of teachers, integration of AWE systems in instructional programs) and research gaps (e.g., lack of studies with experimental control group designs, little consideration of individual learner-related influencing factors). However, they were not able to quantify the effect of AWE feedback on writing.

To this end, however, meta-analytical evidence regarding the effectiveness of computer-based feedback in the context of writing quality is scarce. Initial evidence can be found in the subgroup analyses of the meta-analysis by Graham et al. (2015). In this subgroup analysis, the authors synthesized the effects of computer-based feedback approaches on writing quality, based on $k = 4$ studies. The authors obtained a small to medium ($d = .38$,

95% CI [.17, .59]) significant effect with a large heterogeneity. The restricted number of included studies, however, requires a careful interpretation of the findings. Furthermore, these studies differed from those which were included in the meta-analytic report by Graham et al. (2011). The overall effect over the studies ($k = 4$) in Graham et al. (2011) was not significant, indicating that the effect of computer-based feedback is not robust and further research is needed.

More recently, Fleckenstein, Liebenow and Meyer (2023), Ngo et al. (2022), and Zhai and Ma (2022) performed meta-analyses to further examine the effectiveness of AWE systems. Each of these meta-analyses revealed moderate to large ($g = 0.55$ to $g = 0.86$) positive effects of AWE feedback on learners' overall writing quality. However, besides the fact that these meta-analyses focused only on AWE tools, all of them showed a large heterogeneity across the studies (e.g., writing in first and second language with different emphases), making interpretation of the data difficult. Fleckenstein, Liebenow and Meyer (2023) found some significant moderators of the feedback effect (e.g., educational level, language, time). However, these moderator variables were more related to the intervention itself or the context. To date, it is still unclear what the feedback- and learner-related boundary conditions of computer-based feedback in general (not exclusively AWE systems) are and what factors might explain the heterogeneity found in all of the previous research syntheses.

3.5 Potential Boundary Conditions of Computer-Based Feedback

Based on previous theoretical considerations, it can be argued that the effectiveness of computer-based feedback may depend on distinct system-related, but also learner-related boundary conditions (see Panadero & Lipnevich, 2022; Shute, 2008; Wisniewski et al., 2020; for characteristics of feedback in general; see Fleckenstein, Liebenow, & Meyer, 2023; Kluger & DeNisi, 1996; Narciss, 2008; Ngo et al., 2022; Patchan et al., 2016; Strobl et al.,

2019, in particular for characteristics of computer-based feedback). In the current meta-analysis, we considered four influential factors which sought to impact the effectiveness of computer-based feedback to improve students' writing quality: specificity of feedback, representation of feedback, text level on which feedback was provided, and writers' individual prior knowledge.

3.5.1 Does the Specificity of Feedback Matter?

Computer-based feedback technologies may have the potential to provide specific information to enhance writing quality (Goodman et al., 2004). Specificity of feedback refers to the level of detail of the feedback information (Goodman et al., 2004; Patchan et al., 2016).

Previous research has shown that specific feedback contributes to enhancements of writing quality (Hattie & Timperley, 2007; Shute, 2008; but also, Wisniewski et al., 2020). To this end generic feedback is perceived as less easy to be implemented and thus may result in lower quality gains (Kluger & DeNisi, 1996; Patchan et al., 2016; Shute, 2008). For instance, Lachner et al. (2017a), compared effects of generic versus specific computer-based feedback on the cohesion of university students' expository texts. They varied the type of feedback (specific concept map feedback vs. specific outline feedback vs. general feedback) and examined its effect on the local and global cohesion of student revisions. The results showed that writers perceived generic feedback as more difficult than specific feedback. Specific feedback also contributed to writing performance (as measured by the level of local cohesion).

3.5.2 Does the Representation Format of Feedback Matter?

Previous computer-based feedback technologies not only differed regarding the specificity of the feedback implementation, but also regarding to the utilized representation. Numerical representations provide information in the form of numbers (e.g., overall scores, grades, percentages, or points achieved; Kellogg et al., 2010; Roscoe, Brandon et al., 2013;

Wilson & Roscoe, 2020). Numerical feedback is often visualized in the form of charts, diagrams, and rating scales. Additionally, there are rather complex embedded formats that represent feedback information directly within the text, for example, by signaling, markings, or highlighting (Burkhart et al., 2020; Moore & MacArthur, 2016; Tsai et al., 2020), allowing for easier integration of feedback into the draft. More recent approaches additionally allow for pictorial representations of a text in which the feedback information is represented, for example, as a graph, in form of a concept map, or using symbols (Lachner et al., 2017a, 2017b; O'Rourke et al., 2011; Pirnay-Dummer & Ifenthaler, 2011; Villalon & Calvo, 2011). However, it is still unclear which representation format is best suitable for providing computer-based feedback to support learners in revising their texts and to enhance their writing quality effectively. In our meta-analysis, we were interested in exploring if it makes a difference how the computer-based feedback is presented. Furthermore, we were interested in whether the combination of multiple representations would additionally contribute to writing quality than mono-representational formats (e.g., Ainsworth, 2006) as different representations can complement, constrain, or relate to each other, enhancing students to process their revisions more deeply and gain a deeper understanding of the feedback (cf. Ainsworth, 2006; Scaife & Rogers, 1996).

For instance, in their experiment, Burkhart et al. (2020) compared effects of different versions of computer-based feedback on writing quality. In two conditions, students received concept map feedback which highlighted cohesion gaps of their texts in addition to their text representation. An amended version, additionally, explicitly signaled text-picture relations between the students' draft and the concept map feedback. The contiguous feedback condition directly highlighted the cohesion gaps within the text representation (i.e., mono-representation), whereas in the no-feedback condition, students did not receive any feedback. The authors found benefits of the mono-representation for enhancing the local cohesion of a text. The amended multiple representation format (concept map and text representation) had

benefits regarding global cohesion. These findings suggest that mono- versus multiple representational feedback differently contributed to students' writing quality.

3.5.3 Does it Matter on Which Level of Text Quality Feedback is Provided?

Besides the design of computer-based feedback, particularly the level of text quality, that is the type of information, which is provided, is considered to be important to enhance students' writing quality (Patchan et al., 2016). Feedback on a *lower hierarchical level* (also called micro-level or low-level feedback; Strobl et al., 2019) focuses on surface features of the text and refers, for example, to grammar, spelling, text length, or word count. Feedback on a *higher hierarchical level* (also called macro-level or high-level feedback; cf. Strobl et al., 2019) refers to the organization and structure, the style, or the cohesion of the text. Commonly, it is assumed that improvements of writing quality may particularly emerge, when higher level feedback is provided (Chanquoy, 2009; Patchan et al., 2016). For instance, Patchan et al. (2016) analyzed the effects of the text-level of computer-mediated peer-comments on writing quality. The authors found that only higher-level comments contributed to writing quality. Given that most computer-based feedback systems rather provide information on a lower text level (see Strobl et al., 2019), it is an open issue, whether the text level may account for writing quality in the context of computer-based feedback.

3.5.4 Does Students' Prior Knowledge Matter?

Additionally, the level of students' prior knowledge may affect the effectiveness of computer-based feedback. Such effects are often explained against the expertise-reversal effect, which postulates that the effectiveness of instructional interventions, such as feedback may depend on the level of prior knowledge (Bromme et al., 2004; Kalyuga, 2007; Kalyuga et al., 2003; Nihalani et al., 2011). Whereas low prior-knowledge students may require additional assistance during learning, instructional support may have negative consequences

for more experienced students due to redundant processing of information (Ngo et al., 2022; Richter et al., 2016).

For instance, Fyfe and Rittle-Johnson (2016) gave elementary students strategy instruction about solving mathematical equivalence problems or not to experimentally induce students' prior knowledge. Afterwards, in a practice phase, the students solved 12 mathematical problems. Additionally, students were either provided with feedback or not. The authors found a knowledge-by-feedback interaction, as only low-prior knowledge students profited from the feedback. Whether these findings may replicate in writing settings, is an open question.

3.6 Overview of the Present Meta-Analysis

The aims of the present study were to examine the effectiveness of computer-based feedback regarding the quality of writing, and analyze potential boundary conditions, based on the previous considerations. In contrast to earlier research syntheses, we focused on formative feedback (excluding summative feedback) and system-generated feedback only. Thus, we did not include computer-mediated feedback (e.g., via peer-review systems) which could likely confound our findings. Additionally, we focused on first language writing.

Previous meta-analyses that examined the effectiveness of computer-based feedback on writing could only include a restricted set of studies (Graham et al., 2015; considering articles from 1975 to 2011) or focused on AWE systems only (Fleckenstein, Liebenow, & Meyer, 2023; Ngo et al., 2022; Zhai & Ma, 2022). Therefore, we aimed to update the meta-analytical knowledge, and additionally investigated potential feedback-related and student-related boundary conditions for the effective use of different computer-based feedback technologies to support writing.

3.7 Method

Based on Borenstein et al. (2009), we conducted the meta-analysis according to the three proposed steps: Identifying relevant studies, coding study characteristics, and computing mean effect sizes and their statistical significance. The meta-analysis was preregistered (see [preregistration](#)) via *preregRS* (J. Schneider et al., 2021).

3.7.1 Identifying Relevant Studies

Using the PICO framework (Population – Intervention – Comparison – Outcome) framework (Fineout-Overholt & Johnston, 2005; Schardt et al., 2007), we developed the following search term: `(computer* OR automat*) AND (writ* AND (argument* OR essay* OR summary OR exposit* OR expla*)) AND (feedback OR evaluat* OR assess* OR scor*) NOT (peer OR medic* OR neural-network OR health* OR care*)`. By integrating the NOT-string, in our search term, we were able to exclude studies, which were not relevant to writing, as feedback, is a broad term, which is also used in computer science, medicine or care work. We nevertheless screened, whether we would erroneously exclude relevant studies by conducting a pilot search, which was not the case in our study. We considered the timespan from January 2003 until December 2020, because around 2003 several tools were either established or got a major upgrade (e.g., MyAccess was established in 2003 and Summary Street was established between 2002 and 2004; Wade-Stein & Kintsch, 2004). To retrieve empirical studies of potential relevance, we performed searches of the following databases: Education Resources Information Center (ERIC; via EBSCOhost), PsychInfo, Web of Science, Google Scholar (first 100 results), PsychArchives, Psycharxiv, and SocArcXiv. The searches yielded in 2,459 hits. We carefully followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses' (PRISMA) framework of 2020 (Moher et al., 2009; Page et al., 2021; see PRISMA Flow Diagram, see Figure 6). First, the duplicates of the search results were removed. Second, the remaining 2,299 articles were

Method

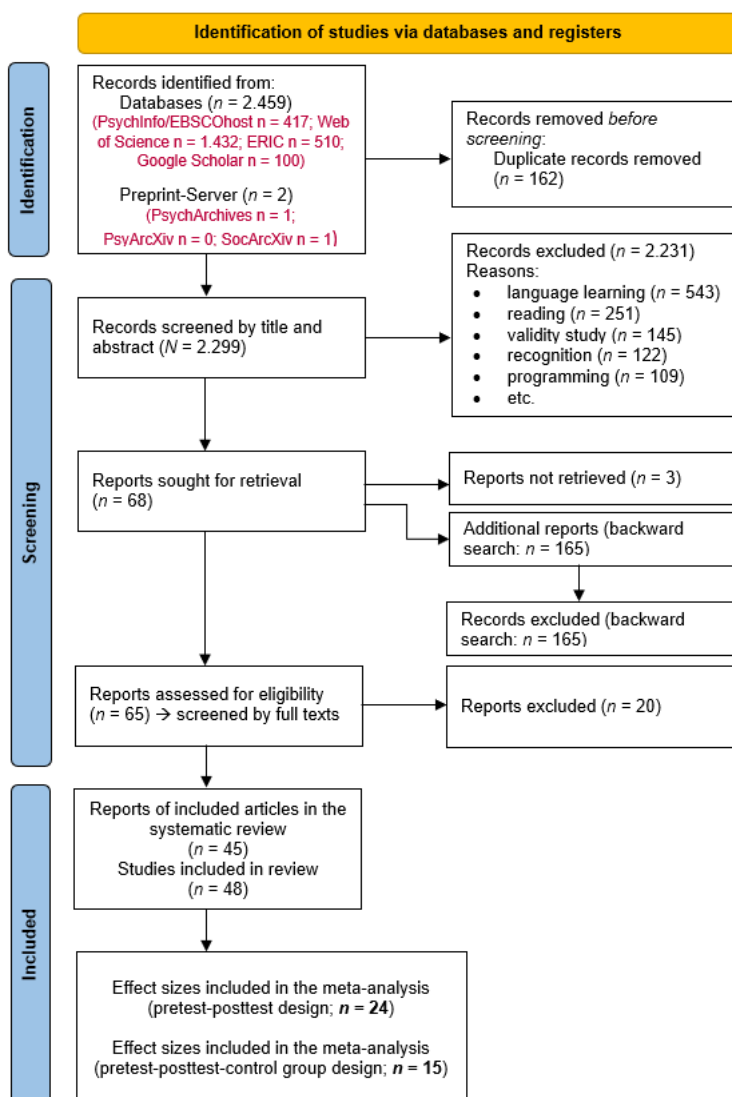
examined by two independent raters to identify studies for inclusion in the systematic review and the meta-analyses. Using the pre-registered inclusion and exclusion criteria (according to the PICOS framework; Fineout-Overholt & Johnston, 2005; Schardt et al., 2007), the articles were first screened by titles and abstracts and then by full texts by two independent raters. We included studies that implemented a feedback tool automatically providing system-generated formative feedback on students' writing quality to be able to exclude a person's opinion, judgement, or expression as a potential influencing factor. Furthermore, the tool should provide feedback on first-language writing drafts to avoid fostering not writing quality, but rather language learning, for example. For the meta-analysis, we only considered studies that followed either a pretest-posttest design or the more rigorous pretest-posttest-control group design to obtain effect sizes and to be able to predict the effectiveness of feedback. We included a broad set of manuscripts to counteract potential effects of publication bias, as long as they were based on empirical research. Articles were excluded when the feedback described in the study was teacher or peer-mediated feedback or when the study was a validation study focusing on the technical functionality of the tool instead of students' learning outcome. All studies were double coded by a scientific collaborator and the first author of this study ($\kappa = 0.94$). After both raters had independently screened all records, the remaining 28 conflicts were resolved through discussion.

Finally, we could derive 24 independent effect sizes for pretest and posttest from 13 of the included articles. Eight of these articles considered a control group. In total, we derived 15 independent effect sizes for studies with a pretest-posttest-control group design (which were also part of the meta-analysis with pretest-posttest design including 24 effects). Thus, we performed a meta-analysis with 24 effect sizes of studies with a pretest-posttest design, and a sub-meta-analysis with the 15 effect sizes of studies with a pretest-posttest-control group design. We deliberately decided against pooling within- and between-effect sizes, as

they result from different metrics (measure of change versus measure of group comparisons), and thus, pooling would likely result in distorted findings (see also Morris & DeShon, 2002).

Figure 6

Procedure According to the PRISMA Framework (Inspired by Page et al., 2021)



3.7.2 Coding

Three independent raters ($\kappa = 0.67$) coded the studies using 33 items collecting information about each study: reference information (e.g., authors, title), sample information (e.g., students' age, educational level), study conduction (e.g., study setting), feedback information and materials (e.g., feedback tool, text genre). For the meta-analyses the final

coding scheme consisted of 19 variables (i.e., authors, scale of the writing quality variable, mean and standard deviation of writing competence in the pretest and posttest for feedback and control condition).

3.7.3 Research Design and Outcome Measures

In the included studies, students wrote a draft (pretest; e.g., an argumentative essay about a given topic) and after receiving feedback on their writing quality (or no feedback in the control condition) they revised the draft (posttest). Therefore, we applied students' writing quality in the pretest as independent variable and students' writing quality in the posttest as dependent variable.

3.7.4 Moderator Variables

We assessed the impact of the following potential boundary conditions on the effectiveness of computer-based feedback in supporting students' writing quality during the revision processes: specificity of feedback (generic vs. specific), representation of feedback (e.g., numeric vs. graphical), level of text quality on which feedback was provided (lower vs. higher level), and students' prior knowledge (see Table 2 for a detailed explanation of the coding of the moderator variables). All moderator variables were coded by three independent raters ($\kappa = 0.73$, substantial according to Fleiss' Kappa). We conducted moderator analyses using the multilevel random effects model.

Table 2*Moderator Variables*

Variable	Description	Scales	Prototypical example of Coding
Text level	Feedback on a lower hierarchical level focuses on surface features of the text (e.g., grammar, spelling, text length, word count). Feedback on a higher hierarchical level is more complex and has a larger influence on writing quality and text comprehension (e.g., organization and structure, style, cohesion).	1 = Low-level only 2 = High-level only 3 = Both lower and higher level	In the study by Roscoe et al. (2015), feedback was provided on different text properties ranging from lower level aspects, e.g., number of words and sentences, to more sophisticated aspects on higher level, e.g., cohesion. Thus, this study was coded with 3 (<i>both lower and higher level</i>).
Graphical feedback representation	Feedback is visualized graphical, i.e., in form of a concept map.	0 = Not graphical 1 = Graphical	Burkhart et al. (2020) provided one group of students feedback in form of a conventional concept map which was coded with 1 as graphical representation.
Numeric feedback representation	Feedback is visualized numeric, e.g., as scores, grades, points achieved, percentages, diagrams, or rating scales.	0 = Not numeric 1 = Numeric	In the study by Kellogg et al. (2010), feedback was provided in form of a holistic score. Thus, this study was coded with 1 (<i>numeric representation of feedback</i>).
Highlighting representation of feedback	Feedback information is directly embedded within the text through signaling, markings, or symbols.	0 = No highlighting 1 = Highlighting was used	In the study by Burkhart et al. (2020), in the signaled concept map condition, signaling was used embedded directly in students' draft and within the concept map. This case was coded with 1 as <i>highlighting</i> was used to represent the feedback information.
Text-based feedback representation	Feedback is represented as text-based comment, hints, messages, or suggestions.	0 = Not text-based 1 = Text-based	In the study by Lachner et al. (2017a), in the outline condition, the feedback information was represented only as keywords why it was coded with 1 (<i>text-based</i>).

Method

Variable	Description	Scales	Prototypical example of Coding
Number of representations	Number of representations describes whether only a single representation format was used to provide the feedback or whether multiple representation formats were combined.	0 = Mono (single) representation 1 = Multiple representations	In the study by Lachner et al. (2017b), feedback regarding local and global cohesion was provided by using a concept map (graphical). Thus, this study was coded with 0 (<i>mono representation</i>). In contrast, in the study by Palermo (2017), feedback was presented in form of scores in bar charts (numeric) combined with written statements (text-based). Thus, this study was coded with 1 (<i>multiple representations</i>).
Feedback specificity	Feedback specificity describes how detailed and comprehensive feedback is provided. In generic feedback, the writing product is considered as a unit (overall score). In specific feedback procedure, feedback is given on different criteria of the text.	1 = Generic (only one generic score or one generic feedback method is provided) 2 = Specific (there are more than one feedback representation and possibly additional suggestions or prompts to help implementing the (localized) feedback are provided)	In the study by Wang et al. (2020), students receive a grade on their first and a grade on their second draft as feedback. Thus, this study was coded with 1 (<i>generic</i>). In contrast, in the study by Wilson and Roscoe (2020), 6 traits of writing quality were scored on a scale from 1 to 5 and students were provided with feedback on each of the 6 traits, indicating strategies for students to improve their draft regarding the specific trait. Thus, this study was coded with 2 (<i>specific</i>).
Prior knowledge	Students' prior knowledge was measured using the pretest measurements.	Because the different studies used different scales to measure writing quality, we transformed students' pretest values into percentages so that we could measure the moderator variable on a scale from 0% to 100%	In the study by Zhu et al. (2017), they used for students' pretest measures a scale from 0 to 10. The mean writing quality was 5.20. We translated the values into percentage ($5.2 \times 100 \div 10$) resulted in 52%.
Setting		1 = Laboratory 2 = Field study	
Educational level		0 = School (including middle and high school) 1 = University	

3.7.5 Computing Effect Sizes and Statistical Analyses

3.7.5.1. *Not Reported Data and Sensitivity Analysis*

A common challenge in meta-analyses is dealing with unavailable statistical parameters from the primary studies (Lipsey & Wilson, 2001). Some studies did not provide the minimally required statistical information to conduct a meta-analysis or provided an imprecise operationalization of writing quality. Accordingly, we had to exclude several studies, as we were not able to gather the missing information. To this end, $n = 24$ (with pretest-posttest design and $n = 15$ with pretest-posttest-control group design) of the 48 identified studies could be included in the meta-analysis, as they provided sufficient detail. We used the means, standard deviations, sample sizes, and pretest-posttest correlations to calculate effect sizes (standardized mean change) for the samples of primary studies via the R-package “metafor” (Viechtbauer, 2010). However, we did not find the pretest-posttest correlation reported in any of the articles. In these cases (single missing statistical information), it is typically suggested to run a sensitivity analysis (Cooper et al., 2019). Sensitivity analyses use a range of plausible values for the missing statistical value, which are then used to continue the data analysis and investigate the variability of results (McKenzie et al., 2022). One of the studies with control group design provided open data (Burkhart et al., 2020). Based on this study we were able to calculate its pretest-posttest correlations (intervention group: $r = .72$; control group: $r = .86$), which we used as an anchor for the sensitivity analysis. We included a broad range of 26 (conservative) plausible values in the sensitivity analyses (cf. Cooper et al., 2019) with correlations from $r = .50$ to $.75$ for the intervention group samples and correlations from $r = .64$ to $.89$ for the control group samples, using steps of $.1$ respectively.

In the results section, we report the mean meta-analytic effect sizes and the according p -values. For a better overview and since the variations do not change the result, the ranges of

the effect sizes and p -values based on the sensitivity analyses are not mentioned in the continuous text but can be found in the according tables (Tables 3–6).

3.7.5.2. Average Effect

For our meta-analysis, we calculated random effect models to account for the heterogeneous design of the included studies. As authors provided multiple comparisons in one paper, we also included the paper-ID as a cluster variable to account for statistical dependencies between effect sizes. We conducted an ANOVA comparing the multilevel random effect model (clustered regarding paper-ID) and the standard model (without cluster variable) to test whether nesting within studies is necessary. The results showed a significant ($p < .01$) preference of the multilevel random effect model over the standard model that indicated that there are substantial differences between papers. Thus, we continued with the random effect model considering paper as cluster variable. As we conducted a meta-analysis on studies with repeated measures (pretest and posttest), we used Becker's (1988) standardized mean change as appropriate effect size measures with qualifying values of $g = .2, .5, .8$ as small, medium, and large effects.

3.7.5.3. Publication Bias

To check for possible publication bias, first, we generated a funnel plot (Borenstein et al., 2009; Sterne et al., 2000) in conjunction with trim-and-fill-analyses which proved to be an effective estimation analysis for determining missing studies (Duval & Tweedie, 2000a, 2000b; Pham et al., 2001). In addition, we used Egger's et al. (1997) regression test to statistically detect the asymmetry of the plot.

3.7.5.4. Used Data Management Tools

Data was gathered and organized using Rayaayn (<https://www.rayyan.ai/>), a systematic review web application and online tool helping researchers to organize and manage

collaboratively the data of systematic reviews and meta-analyses (Ouzzani et al., 2016). Descriptive analyses were carried out using Microsoft Excel. To prepare for statistical analysis, we used SPSS, Version 27. To calculate effect sizes and to conduct the meta-analysis we used the metafor-R-package (Viechtbauer, 2010), which was developed to perform meta-analyses with the R software.

3.8 Results

3.8.1 Sample

Thirteen articles yielding $k = 24$ pretest-posttest comparisons were included in the meta-analysis (see Table 3). The studies were published between 2008 and 2020 as journal articles (77%), dissertations (15%), or conference papers (8%). Most of the studies were conducted in the United States of America (59%), all others in Germany. There were almost as many laboratory studies (46%) as field studies (55%). A total of $N = 2,380$ students participated in the studies with a mean sample size of $n = 99$ (ranging from $n = 20$ to $n = 542$ participants per study). Most of the participants were university students (64%), followed by students in high school (23%), middle school (9%), and a combined sample of middle and high school students (5%). The text genre in which the students were asked to write their drafts was, in most cases, argumentative/persuasive essay writing (46%), followed by writing explanations (41%), scientific argumentations (9%), and summaries (5%).

Eight of these 24 articles yielded $k = 15$ pairwise comparisons of the writing quality of an experimental group that received feedback with a control group that did not receive feedback, with which we conducted a sub-meta-analysis. The experimental group comprised $n = 921$ participants (ranging from $n = 13$ to $n = 287$ among studies; mean sample size of $n = 61$), the control group comprised $n = 908$ participants (ranging from $n = 11$ to $n = 270$ among studies; mean sample size of $n = 61$), resulting in a total sample size of $N = 1,829$ students included in the sub-meta-analysis.

Results

In Table 3, information about all included studies in both meta-analyses are shown in detail.

Table 3

Overview of all Included Studies in the Meta-Analyses

Study	Type of Manuscript	Country	Sample size in Total	Level of Education	Study Design	Sample size per Condition		Feedback tool	Type of Feedback	Text Level	Feedback Representation	Feedback Specificity	Study Setting	Text Genre
						<i>n</i> Experimental Group	<i>n</i> Control Group							
Burkhardt et al., 2020	journal article	Germany	100	university	pre-post-control	25	26	CohViz	correspondence-enhanced concept-map feedback	higher	Highlighting in the text + concept map	specific	laboratory	explanation
	journal article	Germany	100	university	pre-post-control	23	26	CohViz	conventional concept-map feedback	higher	Concept map	specific	laboratory	explanation
	journal article	Germany	100	university	pre-post-control	26	26	CohViz	spatially contiguous feedback	higher	Highlighting in the text + symbols	specific	laboratory	explanation
Frost, 2008	dissertation	USA	31	high school	pre-post-control	23	11	Criterion	AES feedback + teacher instruction	Lower + higher	numeric	generic	field	summary
McCarthy et al., 2019	journal article	USA	60	high school	pre-post	-	-	Writing Pal	Writing Pal feedback (generic score)	Lower + higher	Numeric + text-based	generic	laboratory	(argumentative / persuasive) essay
	journal article	USA	59	high school	pre-post	-	-	Writing Pal	writing Pal feedback + spelling and grammar checking (generic score)	Lower + higher	Numeric + text-based	generic	laboratory	(argumentative / persuasive) essay
Kellogg et al., 2010	journal article	USA	39	university	pre-post-control	20	19	Criterion	intermittent feedback	Lower + higher	numeric	generic	laboratory	(argumentative / persuasive) essay
	journal article	USA	20	university	pre-post-control	20	19	Criterion	continuous feedback	Lower + higher	numeric	generic	laboratory	(argumentative / persuasive) essay
Lachner et al., 2017a	journal article	Germany	96	university	pre-post	-	-	CohViz	general feedback	higher	numeric	specific	field	explanation
	journal article	Germany	77	university	pre-post	-	-	CohViz	concept map feedback	higher	Concept map	specific	field	explanation
	journal article	Germany	78	university	pre-post	-	-	CohViz	outline feedback	higher	text	specific	field	explanation

Results

Study	Type of Manuscript	Country	Sample size in Total	Level of Education	Study Design	Sample size per Condition		Feedback tool	Type of Feedback	Text Level	Feedback Representation	Feedback Specificity	Study Setting	Text Genre
						<i>n</i> Experimental Group	<i>n</i> Control Group							
Lachner et al. (2017b), Study 2	journal article	Germany	42	university	pre-post-control	21	21	CohViz	concept map feedback	higher	Concept map	specific	laboratory	explanation
Lachner et al. (2017b), Study 3	journal article	Germany	27	university	pre-post-control	13	14	CohViz	concept map feedback	higher	Concept map	specific	laboratory	explanation
Lachner & Neuburg, 2019	journal article	Germany	61	university	pre-post-control	31	30	CohViz	concept map feedback with signaling	higher	Highlighting in the text + concept map	specific	laboratory	explanation
Palermo, 2017	dissertation	USA	542	middle school	pre-post-control	272	270	NCWrite	feedback + traditional instruction	lower	Numeric + text-based	specific	field	(argumentative / persuasive) essay
	dissertation	USA	557	middle school	pre-post-control	287	270	NCWrite	feedback + self-regulated strategy development	lower	Numeric + text-based	specific	field	(argumentative / persuasive) essay
Roscoe et al., 2013	conference paper	USA	65	high school	pre-post-control	33	32	Writing Pal	Writing-Pal condition	Lower + higher	Numeric + text-based	generic	field	(argumentative / persuasive) essay
Wang et al., 2020	journal article	USA	143	middle school	pre-post			eRevise	AWE system eRevise	lower	numeric	generic	field	(argumentative / persuasive) essay
Weston-Sementelli et al., 2016	journal article	USA	89	university	pre-post-control	41	48	Writing Pal	iSTART	Lower + higher	Numeric + text-based	generic	field	(argumentative / persuasive) essay
	journal article	USA	41	university	pre-post-control	41	48	Writing Pal	Writing-Pal	Lower + higher	Numeric + text-based	generic	field	(argumentative / persuasive) essay
	journal article	USA	45	university	pre-post-control	45	48	Writing Pal	iSTART + Writing-Pal	Lower + higher	Numeric + text-based	generic	field	(argumentative / persuasive) essay
Zhu et al., 2020	journal article	USA	131	high + middle school	pre-post	-	-	c-rater ML	generic	lower	Numeric + text-based + scales	specific	field	scientific argumentation
	journal article	USA	203	high + middle school	pre-post	-	-	c-rater ML	contextualized	higher	Numeric + text-based + scales	specific	field	scientific argumentation
Zhu et al., 2017	journal article	USA	141	high school	pre-post	-	-	c-rater ML	automated scoring feedback	higher	Numeric + text-based	specific	field	scientific argumentation

3.8.2 Publication Bias

As might be expected with large heterogeneity between studies, publication bias revealed significance. The funnel plot of the pretest-posttest meta-analysis shows that the effect sizes are not systematically distributed around the average effect, but there was a gap at the bottom (particularly on the left side) of the plot. The gap indicates that small (nonsignificant) effect sizes of small studies are missing, suggesting that there may be evidence of a possible publication bias. Egger's regression test (Egger et al., 1997) revealed an asymmetry of the funnel plot, $z = 4.1485$, $p < .01$, $b = -.1501$. However, the trim-and-fill analysis did not include any study but adjusted the center of symmetry. Accordingly, there was a change of the effect size due to the trim-and-fill method resulting in a medium to large significant effect of $g = .68$, $p < .01$, CI 95% [.3787, .9887] (Egger et al., 1997; Sterne et al., 2000).

For the pretest-posttest-control group meta-analysis, there was a different pattern: The funnel plot was relatively symmetrical with a gap at the bottom (right side) of the plot, suggesting that large effect sizes of small studies are missing. However, Egger's regression test (Egger et al., 1997) revealed an asymmetry of the funnel plot, $z = -2.3505$, $p = .019$, $b = .7859$. Performing the trim-and-fill analysis, two studies were included on the right side and there was a slightly change of the effect, $g = .37$, $p = .004$, CI 95% [.1185, .6294].

In both funnel plots, most studies are shown in the upper range, whereas small studies (with small effects in the pretest-posttest meta-analysis and large effects in the pretest-posttest-control group meta-analysis) are rather absent. This suggests that a so-called small-study effect is present, where the results of smaller studies differ consistently from the larger studies (Schwarzer et al., 2015; Sterne et al., 2000), resulting in an asymmetry. In conclusion, the analyses suggest that the few studies which could be integrated in the present meta-analyses might be report biased. However, all these analyses on publication bias should be treated with caution due to the low number of included studies (Jackson & Turner, 2017). In addition, it should be noted that the asymmetries did probably not result from publication

bias but might have other reasons (Egger et al., 1997; Schwarzer et al., 2015). Other possibilities that occur asymmetry of the funnel plot could be unbalanced existence of the characteristics in the population (i.e., the population may not be evenly distributed and therefore we should not use an evenly distributed population as the basis for the comparison), small studies showed different (often larger) effects than studies with a larger sample size (small-study effect; Sterne et al., 2000), artefactual correlations of statistical estimates with their standard errors, or only studies with large or significant treatment results were reported (selective/outcome reporting bias; Chan & Altman, 2005).

3.8.3 Overall Effect Size for Writing Quality

The meta-analysis including 24 independent studies with pretest-posttest design ($N = 2,380$ participants) revealed a standardized mean change of $g = .53$ with $p = .013$, indicating that students improved their writing quality significantly (medium effect) from pretest to posttest. Table 4 shows an overview of the study-wise effect sizes¹ and confidence intervals.

For the pretest-posttest-control group meta-analysis, we included 15 of the studies (involving 1,829 participants) of the meta-analysis for the pretest-posttest designs only which additionally included a control group. In contrast to the first meta-analysis about the studies with pretest-posttest-design only, under consideration of a control group, we found smaller effects which were not significant: Students in the feedback condition did not improve their writing quality more from pretest to posttest compared to students in the control condition, $g = .27$, $p = .084$. Table 5 shows an overview of the study-wise effect sizes² and confidence intervals.

¹ Table 4 shows the ranges of the effect size, its standard error and p -value depending on the pretest-posttest correlation within the scope of the sensitivity analysis. The results showed that the variations of the effect size g are minimal (only from the third decimal place) and do not change the result, which is why we can assume a robust effect and a precise estimate of the correlation.

² Regarding the effects of the studies with pretest-posttest-control group design, the variations of g were larger, depending on the assumed pretest-posttest correlation of our sensitivity simulation (see Table 5). This finding suggests that the effect is not very robust across all 26 meta-analyses but depends on the correlation. The result should therefore be treated with caution.

Table 4

Overview of Effect Sizes and Confidence Intervals of all Included Studies in the Meta-Analysis of Studies with Pretest-Posttest Design ($k = 24$)

Study	Number of participants n	Effect size g	SE	95% CI	
				Lower	Upper
Burkhart et al., 2020	51	1.47	.270 (.251-.288)	.9395 (.9040-.9766)	1.9990 (1.9619-2.0345)
Burkhart et al., 2020	49	.72	.209 (.181-.234)	.3085 (.2593-.3619)	1.1266 (1.0733-1.1759)
Burkhart et al., 2020	52	.61	.189 (.162-.213)	.2358 (.1878-.2882)	.9762 (.9238-1.0242)
Frost, 2008	34	-.54	.197 (.167-.223)	-.9235 (-.9755--.8665)	-.1530 (-.2101--.1010)
Kellogg et al., 2010	39	2.69	.467 (.453-.480)	1.7725 (1.7465-1.7990)	3.6030 (3.5765-3.6290)
Kellogg et al., 2010	39	2.95	.505 (.492-.517)	1.9590 (1.9349-1.9835)	3.9376 (3.9131-3.9617)
Lachner et al., 2017a	96	.13	.088 (.073-.102)	-.0428 (-.0704--0.0121)	.3039 (.2732-.3314)
Lachner et al., 2017a	77	.13	.099 (.081-.114)	-.0616 (-.0923--0.0273)	.3256 (.2913-.3563)
Lachner et al., 2017a	78	.19	.099 (.082-.114)	-.0031 (-.0335-.0307)	.3840 (.3502-.4144)
Lachner et al., 2017b (Study 2)	42	1.43	.290 (.269-.310)	.8598 (.82044-.9010)	1.9969 (1.9557-2.0362)
Lachner et al., 2017b (Study 3)	27	.70	.276 (.240-.310)	.1609 (.0951-.2323)	1.2430 (1.1716-1.3088)
Lachner & Neuburg, 2019	61	1.04	.203 (.183-.223)	.6379 (.6002-.6781)	1.4354 (1.3952-1.4731)
McCarthy et al., 2019	60	.22	.113 (.094-.131)	.0027 (-0.0318-.0410)	.4461 (.4077-.4806)
McCarthy et al., 2019	59	.22	.114 (.094-.132)	-.0074 (-0.0422-.0313)	.4392 (.4005-.4740)
Palermo, 2017	542	.44	.056 (.047-.064)	.3333 (.3178-.3504)	.5513 (.5342-.5669)
Palermo, 2017	557	.97	.065 (.058-.072)	.8436 (.8309-.8572)	1.0989 (1.0853-1.1116)
Roscoe et al., 2013	65	.14	.151 (.124-.175)	-.1564 (-.2034--0.1041)	.4354 (.3831-.4823)
Wang et al., 2020	143	.15	.073 (.060-.084)	.0035 (-0.0190-.0287)	.2880 (.2629-.3105)
Weston-Sementelli et al., 2016	89	-.18	.136 (.112-.157)	-.4507 (-.4926--.4040)	.0828 (.0361-.1247)
Weston-Sementelli et al., 2016	89	.04	.135 (.111-.156)	-.2261 (-.2685--.1789)	.3016 (.2543-.3439)
Weston-Sementelli et al., 2016	93	.65	.146 (.126-.164).	.3654 (.3295-.4046)	.9368 (.8977-.9728)
Zhu et al., 2020	131	1.87	.138 (.131-.145)	1.6023 (1.5890-.1.6160)	2.1440 (2.1303-2.1574)
Zhu et al., 2020	203	1.39	.092 (.085-.010)	1.2064 (1.1935-1.2199)	1.5659 (1.5524-1.5788)
Zhu et al., 2017	141	.35	.076 (.063-.087)	.2036 (.1816-.2280)	.4998 (.4754-.5218)

Results

Note. g is the effect size of Becker's (1988) standardized mean change from students' first draft (pretest) to their revised draft (posttest); the effect size did not vary due to different pretest-posttest correlation. SE is the standard error of the effect size. 95% CI is a 95% confidence interval with its lower and upper limit. The table shows the mean effect size, its standard error, confidence interval, and the corresponding p -value across the 26 meta-analyses based on the mean correlation according to the sensitivity analysis procedure. The range of the respective values of the 26 meta-analyses with the smallest and largest correlation is reported in parentheses. Effect size does not vary due to different pretest-posttest correlation, hence only one effect size value per study is reported.

Table 5

Effect Sizes and Confidence Intervals of all Included Studies in the Meta-Analysis of Studies with Pretest-Posttest-Control Group-Design (k = 15)

Study	Number of participants <i>n</i>	Effect size <i>g</i>	<i>SE</i>	95% CI	
				<i>LL</i> mean (min-max)	<i>UL</i> mean (min-max)
Burkhart et al., 2020	51	.68	.321 (.289-.350)	.05283 (-.0054-.1148)	1.3100 (1.2480-1.3683)
Burkhart et al., 2020	49	-.07	.271 (.231-.307)	-.6009 (-.6722--.5229)	.4604 (.3824-.5316)
Burkhart et al., 2020	52	-.18	.256 (.216-.292)	-.6831 (-.7538--.6053)	.3195 (.2417-.3901)
Frost, 2008	34	-.06	.281 (.216-.337)	-.6075 (-.7182--.4802)	.4939 (.3665-.6045)
Kellogg et al., 2010	39	-1.07	.784 (.768-.901)	-2.6118 (-2.6436--2.5795)	.4629 (.4306-.4948)
Kellogg et al., 2010	39	-.81	.807 (.791-.823)	-2.3964 (-2.4273843--2.3651)	.7687 (.7373-.7996)
Lachner et al., 2017b (Study 2)	42	.87	.337 (.300-.371)	.2120 (.1449-.2838)	1.5331 (1.4613-1.6002)
Lachner et al., 2017b (Study 3)	27	.49	.333 (.273-.386)	-.1651 (-.2693--.0488)	1.1394 (1.0231-1.2436)
Lachner & Neuburg, 2019	61	.38	.253 (.219-.284)	-.1192 (-.1804--.0527)	.8730 (.8066-.9342)
Palermo, 2017	542	.64	.070 (.056-.082)	.5040 (.4794-.5318)	.7774 (.7496-.8020)
Palermo, 2017	557	1.17	.078 (.065-.089)	1.0175 (.9958-1.0415)	1.3217 (1.2978-1.3435)
Roscoe et al., 2013	65	-.02	.194 (.151-.231)	-.4029 (-.4765--.3186)	.3567 (.2724-.4303)
Weston-Sementelli et al., 2016	89	-.07	.168 (.132-.200)	-.4025 (-.4649--.3312)	.2559 (.1847-.3184)
Weston-Sementelli et al., 2016	89	.15	.167 (.130-.199)	-.1784 (-.2413--.1065)	.4753 (.4034-.5381)
Weston-Sementelli et al., 2016	93	.76	.176 (.143-.205)	.4170 (.3600-.4810)	1.1067 (1.0427-1.1638)

Note. *g* is the effect size of Becker's (1988) standardized mean change from students' first draft (pretest) to their revised draft (posttest) in comparison with a control group, *SE* is the standard error of the effect size, 95% CI is a 95% confidence interval with its lower and upper limit. The table shows the mean effect size, its standard error, confidence interval, and the corresponding *p*-value across the 26 meta-analyses based on the mean correlation according to the sensitivity analysis procedure. The range of the respective values of the 26 meta-analyses with

Results

the smallest and largest correlation is reported in parentheses. Effect size does not vary due to different pretest-posttest correlation, hence only one effect size value per study is reported.

3.8.4 Heterogeneity Analyses

There was statistically observable heterogeneity in the set of effect sizes in both meta-analyses, $Q(23) = 473.68$, $p < .001$, $I^2 = 98\%$ (ranging from 97% to 98%) for the meta-analysis for the pretest-posttest design including 24 studies, and $Q(14) = 117.38$, $p < .001$, $I^2 = 82\%$ (ranging from 77% to 87%) for the meta-analysis with 15 studies with pretest-posttest-control group design. The I^2 statistic indicated that on average 98% in the pretest-posttest meta-analysis and 82% in the meta-analysis for the pretest-posttest-control group designs of the detected variation could be related to true variation among studies, which refers to high heterogeneity (Higgins et al., 2003). To explore possible reasons of this heterogeneity, we conducted several separate moderator analyses using the multi-level random-effects model.

3.8.5 Moderator Analyses

For the meta-analysis about 24 effect sizes from studies with pretest-posttest design, several moderators approached significance. Regarding feedback representation, the moderation analyses showed that graphical and highlighting feedback improved students writing quality positively whereas text-based feedback decreased the writing quality (see Table 6). Numeric feedback had no impact on students' writing quality. Regarding text level, the results indicated that students who received low-level feedback only improved their writing significantly, whereas when students received both low- and high-level feedback their writing quality decreased. High-level feedback only made no difference. Regarding feedback specificity, there was a medium-to-large positive average effect on students' writing quality indicating that more specific feedback led to larger improvements in student writing quality than generic (non-specific) feedback.

Table 6*Results of Moderator Analysis with Studies with Pretest-Posttest Design*

Moderator variables	<i>g</i>	<i>SE</i>	95% CI		<i>p</i>
			Lower	Upper	
Text level^a	-.51 (-.53--.50)	.125 (.116-.132)	-.758 (-.762--.756)	-.270 (-.307--.238)	.000 (.000-.000)
Low text level^b	.45 (.45--.46)	.160 (.151-.168)	.141 (.121-.161)	.767 (.753-.779)	.005 (.002-.007)
High text level ^b	-.04 (-.06--.00)	.130 (.120-.138)	-.292 (-.335--.240)	.216 (.208-.231)	.775 (.647-.968)
Both text levels^b	-.67 (-.70--.64)	.208 (.188-.226)	-1.076 (-1.084--1.064)	-.260 (-.326--.200)	.002 (.000-.004)
Feedback representation					
Graphical^b	.34 (.31-.36)	.094 (.080-.107)	.156 (.153-.158)	.526 (.450-.572)	.000 (.000-.001)
Numeric ^b	-.19 (-.20--.17)	.113 (.095-.129)	-.411 (-.453--.360)	.032 (.013-.051)	.092 (.068-.119)
Highlighting^b	.52 (.51-.54)	.171 (.151-.190)	.188 (.162-.212)	.860 (.805-.908)	.002 (.001-.0048)
Text-based^b	-.77 (-.77--.76)	.153 (.137-.168)	-1.069 (-1.093--1.039)	-.468 (-.504--.434)	.000 (.000-.000)
Number of representations ^c	.12 (.11-.13)	.204 (.202-.243)	-.316 (-.344--.287)	.564 (.506-.610)	.580 (.577-.590)
Feedback specificity^d	.68 (.66-.71)	.206 (.187-.223)	.281 (.226--.342)	1.089 (1.075-1.101)	.001 (.000-.003)
Prior knowledge ^e	-.01 (-.01--.01)	.013 (.011-.014)	-.033 (-.037--.029)	.156 (.015-.017)	.477 (.456-.526)
Setting^f	-.82 (-.82--.80)	.198 (.179-.215)	-1.204 (-1.244--1.155)	-.427 (-.452--.401)	.000 (.000-.000)
Educational level^g	.73 (.71-.74)	.206 (.187-.223)	.325 (.277-.378)	1.133 (1.110-1.152)	.000 (.000-.000)

Note. Number of studies and effects = 24, total $N = 2,380$ students, g is the effect size of Becker's (1988) standardized mean change from students' first draft (pretest) to their revised draft (posttest), SE is the standard error of the effect size, 95% CI is a 95% confidence interval with its lower and upper limit. The table shows the mean effect size, its standard error, confidence interval, and the corresponding p -value

across the 26 meta-analyses based on the mean correlation according to the sensitivity analysis procedure. The range of the respective values of the 26 meta-analyses with the smallest and largest correlation is reported in parentheses.

^a 1 = feedback that addressed lower level of text quality only, 2 = feedback that addressed higher level of text quality only, 3 = feedback that addressed both lower and higher level of text quality. ^b 0 = no, 1 = yes. ^c 0 = mono (one single representation format), 1 = multiple representation formats combined. ^d 1 = generic, 2 = specific. ^e Students' writing quality of their first draft (pretest) on a scale from 0% to 100%; it should be noted that the effect shown in the table is not a small decrease in knowledge, but *g* describes the size and direction of the moderation effect. ^f 1 = laboratory, 2 = field study. ^g 0 = school (including middle and high school), 1 = university.

Results

For the meta-analysis about 15 effect sizes from pretest-posttest-control group design, three moderators revealed significance. Regarding text level, we found that similar to the findings of the pretest-posttest meta-analysis, that only low-level feedback led to higher student writing quality, whereas high-level feedback only or feedback on both lower and higher-level decreased students' writing quality (see Table 7). Regarding feedback specificity, there was a medium positive average on students' writing quality suggesting that specific feedback resulted in higher writing quality compared to generic feedback. Regarding prior knowledge, there was a very small negative average effect on students' writing quality indicating that students with higher prior knowledge profited less from the feedback compared to students with lower prior knowledge.

The effects for the other moderator variables were not significant (see Table 7).

Table 7*Results of Moderator Analysis with Studies with Pretest-Posttest-Control Group Design*

Moderator variables	<i>g</i>	<i>SE</i>	95% CI		<i>p</i>
			Lower	Upper	
Text level^a	-.52 (-.52--.52)	.090 (.077-.100)	-.717 (-.696--.672)	-.345 (-.371--.323)	.000 (.000-.000)
Low text level^b	.55 (.54-.56)	.103 (.085-.119)	.351 (.330-.376)	.755 (.711-.796)	.000 (.000-.000)
High text level^b	-.48 (-.49--.46)	.102 (.084-.117)	-.679 (-.693--.660)	-.280 (-.330--.234)	.000 (.000-.000)
Both text levels^b	-.53 (-.54--.52)	.267 (.254-.278)	-1.057 (-1.088--1.019)	-.012 (-.025-.001)	.045 (.040-.0497)
Feedback representation ^c					
Graphical ^b	.35 (.35-.36)	.248 (.222-.270)	-.131 (-.179--.079)	.840 (.790-.880)	.152 (.109-.192)
Numeric ^b	-.17 (-.18--.17)	.350 (.337-.360)	-.860 (-.872--.843)	.510 (.479-.539)	.617 (.589-.644)
Highlighting ^b	.13 (.13-.13)	.268 (.239-.293)	-.393 (-.445--.336)	.658 (.602-.702)	.619 (.579-.661)
Text-based ^b	.25 (.24-.27)	.323 (.316-.327)	-.378 (-.379--.375)	.887 (.861-.909)	.430 (.415-.446)
Number of representations ^c	.26 (.24-.28)	.244 (.220-.264)	-.215 (-.235--.194)	.741 (.668-.801)	.281 (.280-.284)
Feedback specificity^d	.53 (.52-.54)	.267 (.254-.278)	.012 (.001-.025)	1.057 (1.019-1.088)	.045 (.040-.0497)
Prior knowledge^e	-.05 (-.05--.04)	.021 (.019-.023)	-.091 (-.098--.081)	-.007 (-.007--.007)	.021 (.021-.023)
Setting ^f	.12 (.10-.14)	.336 (.329-.342)	-.543 (-.548--.536)	.775 (.740-.806)	.730 (.693-.770)
Educational level ^g	-.11 (-.14--.09)	.339 (.331-.344)	-.779 (-.812--.741)	.549 (.538-.557)	.736 (.690-.781)

Note. Number of studies and effects = 15, total $N = 1,829$ students, g is the effect size of Becker's (1988) standardized mean change from students' first draft (pretest) to their revised draft (posttest), SE is the standard error of the effect size, 95% CI is a 95% confidence interval with its lower and upper limit. The table shows the mean effect size, its standard error, confidence interval, and the corresponding p -value

Results

across the 26 meta-analyses based on the mean correlation according to the sensitivity analysis procedure. The range of the respective values of the 26 meta-analyses with the smallest and largest correlation is reported in parentheses. Significant results are highlighted in bold letters.

^a 1 = feedback that addressed lower level of text quality only, 2 = feedback that addressed higher level of text quality only, 3 = feedback that addressed both lower and higher level of text quality. ^b 0 = no, 1 = yes. ^c 0 = mono (one single representation format), 1 = multiple representation formats combined. ^d 1 = generic, 2 = specific. ^e Students' writing quality of their first draft (pretest) on a scale from 0% to 100%; it should be noted that the effect shown in the table is not a small decrease in knowledge, but *g* describes the size and direction of the moderation effect. ^f 1 = laboratory, 2 = field study. ^g 0 = school (including middle and high school), 1 = university.

3.9 Discussion

3.9.1 Summary of Evidence

The main aim of the present meta-analysis was to investigate the effectiveness of system-generated feedback (compared to no feedback) to support students' writing quality, and to investigate potential feedback-related and student-related boundary conditions of the effectiveness of computer-based feedback. We found a medium effect of studies with a pretest-posttest design. However, the effect was no longer significant compared to a control group (studies with pretest-posttest-control group design). Thus, overall, we conclude that computer-based feedback is not effective per se.

The higher effects of the pretest-posttest designs can be assumed, as simple pretest-posttest designs rather measure change instead of the effect of an intervention, and also may be prone to other confounds, such as maturation effects (see also Morris, 2008). The effect sizes are in concordance with the meta-analytical findings by Graham et al. (2015), who also documented small to medium effects of computer-based studies, based on a restricted sample of four studies. Additionally, we also obtained a high heterogeneity among the included studies (see also Ngo et al., 2022; Nunes et al., 2021; Strobl et al., 2019; for comparable findings). Part of the variance could be explained by system-related factors: As a consistent pattern across both meta-analyses, we identified the level of text quality and the specificity of feedback as moderators. Low-level feedback increased writing quality more than high-level feedback or feedback that provided both low- and high-level feedback. This finding is surprising, as previous studies suggested that particularly high-level feedback should contribute to writing quality (Crossley & McNamara, 2016; McNamara et al., 2013; Patchan et al., 2016; Strobl et al., 2019). One explanation could be that the implementation of high-level feedback is presuppositional and requires, for instance, numerical or graphical understanding, high levels of self-regulation and therefore additional assistance during the revision process (Lachner et al., 2017b; McNamara et al., 2015). Therefore, the

implementation of low-level feedback could have been less difficult (Berninger & Swanson, 1994; Chanquoy, 2001; Chenoweth & Hayes, 2001; Hacker, 1994; Hayes & Olinghouse, 2015) which could have resulted in a spill-over in high-level implementations and contribute to writing quality (see also Lachner et al., 2017b, for related discussions). However, this interpretation is highly speculative, and requires additional studies, which directly test this hypothesis.

Another interesting finding was that the specificity moderated the effectiveness of computer-based feedback as specific feedback was more effective than general feedback. This finding is consistent with previous research (e.g., Kuklick & Lindner, 2023; Lachner et al., 2017a; Mertens et al., 2022; Shute, 2008), and corroborates the need for specific information to implement distinct revision activities to improve writing quality. What constitutes specificity in computer-based feedback, however, goes beyond the scope of this study, as our moderation analyses only allowed to approach the quality of feedback in a coarse-grained manner. Given that our moderation analyses resulted from cross-study comparisons, we want to clarify that these moderation analyses cannot be interpreted in a causal manner.

Regarding the student-related factors, we investigated students' prior knowledge as potential moderator. In line with previous findings (Fyfe & Rittle-Johnson, 2016; Mertens et al., 2022; Ngo et al., 2022), we found that students with high prior knowledge profited less from the feedback than students with lower prior knowledge. A potential explanation could be the expertise-reversal in the effectiveness of feedback depending on students' level of knowledge (Kalyuga, 2007; Kalyuga et al., 2003; Kalyuga & Renkl, 2010). That means that feedback that helps less experienced students may inhibit students with higher prior knowledge improving their writing because they retrieve and process redundant information they already know and so impairs their available cognitive capacity. Additional research is

needed to investigate how feedback could be implemented to support students with different levels of prior knowledge effectively.

3.9.2 What are the Theoretical and Practical Contributions of our Meta-Analysis?

Our main findings showed that it is not only about the feedback that matters, but particularly about its boundary conditions (feedback-related as well as student-related). The extent to which computer-based feedback is embedded remains a blind spot. Orchestration could not be investigated in the meta-analysis due to lack of information or transparency in the primary studies, highlighting the need to explore this further in future research.

The small effect sizes of computer-based feedback document that, compared to person-generated feedback by instructors, tutors or peers, computer-based feedback still considerably fell short of expectations, given that computer-based feedback requires distinct, long-term, multi-professional resources during development. In addition, in ill-defined domains such as writing, the effects can be assumed to be smaller, and more fragile than in other, more well-defined domains, such as STEM (Sibley et al., 2023). Large language models, such as GPT or BERT (Scheurer et al., 2022; Tsai et al., 2020; Wu et al., 2022) as well as technical advancements could help provide more sophisticated assessments and feedback approaches, which could additionally contribute to writing quality. More research, however, is necessary to investigate whether and how large language models can be implemented during writing. Another interesting finding for moving the field forward was the restricted number of empirical studies which tapped into the effectiveness of computer-based feedback to foster expository writing (see also Graham et al., 2015). Therefore, we see the need in addition to the development of more advanced computer-based feedback approaches to additionally investigate the effectiveness of the developed technologies with experimental methods to test for causality. These empirical studies would help to gain a better understanding of the conditions of computer-based feedback and generalize our

findings. Last, our findings also have implications for educational practice, as they show that computer-based feedback can be used as a supplement for teachers to foster writing under consideration of the boundary conditions. It can be seen as a general guideline to provide feedback specifically (Mertens et al., 2022; Shute, 2008). Low-level feedback is easy to implement and improves writing quality for less as well as more experienced writers. However, as previous research showed that feedback on higher-level has a larger impact on text quality (Crossley & McNamara, 2016; McNamara et al., 2013; Strobl et al., 2019), we suggest to not only provide low level feedback, but also provide specific high-level feedback, for example by offering additional support to enable students to understand, structure, and process the feedback (Mertens et al., 2022). Additionally, adaptive learning environments offer potential to provide feedback depending on students' prior knowledge (Kalyuga & Renkl, 2010).

However, there is a lack of field-oriented studies that explore, among other questions, whether it makes a difference whether computer-based feedback is more effective in the classroom or in individual learning sessions. In addition, we see the need to further investigate how computer-based feedback should be orchestrated.

3.9.3 Limitations

A limitation of this meta-analysis is that we only investigated three system-related and one student-related influencing factor. However, the literature and previous research suggests that there are many more influencing factors (e.g., according to the writing model by Hayes, 2012: environment, motivation, schemes; see also Fleckenstein, Liebenow, & Meyer, 2023; Kluger & DeNisi, 1996; Narciss, 2008; Panadero & Lipnevich, 2022) that should be investigated. In our meta-analysis, first indications emerged that, for instance, cognitive load or mental effort might play a role in the context of more complex feedback or feedback at both lower and higher text levels and also in the context of nonspecific feedback.

Another limitation is that we focused on expository writing in first language, and we cannot generalize them on other text genres or foreign language learning. However, previous research showed that feedback tools aiming to foster second language learning provided feedback targeted rather hierarchically lower level of text quality (e.g., spelling and grammar, length, word choice), whereas other feedback tools focused more on content and structure of students' texts (e.g., overall organization of ideas, cohesion; Strobl et al., 2019). The literature also shows indicators that there are genre-related differences regarding the feedback tools. For example, feedback on academic writing focus more on short and concise sentences and a clear structure of argumentation, whereas feedback on summaries focuses more on the appropriateness and completeness of the content, or feedback tools aiming to foster the writing process focus more on self-regulation skills or the use of writing strategies (Nesi & Gardner, 2012; Strobl et al., 2019).

As a final limitation, we have to mention that the sensitivity analysis showed lack of robustness of some effects, and they have to be interpreted with caution. Future research is needed to further increase the validity of our results. However, this requires more open science and transparency of the primary studies. As a first step in this direction, we set ourselves high methodological standards.

3.9.4 Conclusion

The goal of this meta-analysis was to investigate the effectiveness of computer-based feedback to support students in expository writing with aggregate data. Furthermore, we aimed at identifying potential boundary conditions to complement the existing research and gain new insights for practitioners as well as for researchers into computer-based formative feedback to improve expository writing automatically. Overall, the meta-analysis presented in this paper provides important evidence that specific computer-based feedback particularly

Discussion

on lower text level improved students' writing quality, especially for less experienced writers.

4 THE MORE, THE BETTER?

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>

This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. The Version of Record of this manuscript has been published and is available in the journal *Learning and Instruction* in 2024 (and online available since October 27, 2023).

Structured Abstract

Background: Strategy instruction and formative feedback are powerful instructional strategies to support learning. However, combining them may not necessarily be effective. Previous research on the combination of strategy instruction and feedback produced both additive and reducing effects on learning. It is an open question whether and under which conditions combining strategy instruction and feedback enhances students' learning.

Aims: We investigated potential synergy and interaction effects of combining strategy instruction and feedback on learning in physics, the underlying mechanisms, and their generalizability (elaboration level of feedback, instructional sequence).

Sample: We conducted three online experiments ($N_1 = 437$, $N_2 = 310$, $N_3 = 166$) with university students.

Method: In Experiment 1 and 2, in the learning phase, participants read a strategy instruction or no instruction. In a subsequent practice phase, they solved troubleshooting tasks on electric circuits and received either feedback or no feedback. In Experiment 3, students were randomly assigned to one of three conditions (no intervention, learning phase first, practice phase first). We used the same material for all experiments.

Results: In Experiment 1, we found no additive combination effects of strategy instruction and corrective feedback. Surprisingly, for far transfer, corrective feedback reduced the effect of prior instruction. Contrarily, in Experiment 2, we obtained additive effects by combining strategy instruction with elaborated feedback. Experiment 3 showed that these additive findings were irrespective of the instruction-feedback-sequence.

Conclusions: Our findings indicate that the elaboration level of feedback determines whether combining instruction and feedback produces additive or reducing effects.

Keywords: strategy instruction, example-based learning, feedback, arousal, physics

4.1 Introduction

In the course of the digital transformation, online learning environments have become paramount in educational settings. Two frequently applied instructional interventions to support students' learning in online environments are strategy instruction and formative feedback in subsequent practice phases. For instance, in the domain of electric circuits troubleshooting, strategy instruction is given by providing a conceptional explanation about Ohm's law together with worked examples of how to solve electric circuit problems to help students gain a robust understanding of a problem solving strategy with its different solution steps (Renkl, 2014; van Gog et al., 2006, 2011; van Harsel et al., 2020). Additional practice tasks in combination with feedback should help apply and transfer these strategies to different, yet related problem-situations (e.g., Fyfe & Rittle-Johnson, 2016; Huitt et al., 2009). From a pragmatic stance, it makes sense to combine strategy instruction and feedback to maximize learning which is also reflected in state-of-the-art instructional design models (e.g., Huitt et al., 2009; van Merriënboer et al., 2002). Previous research, which directly examined combinations of strategy instruction and feedback, however, is scarce and, if available, produced mixed findings as the combination of strategy instruction and feedback yielded both additive (e.g., Salden et al., 2010) and reducing effects on learning (e.g., Fyfe & Rittle-Johnson, 2016).

In this paper, we present data from three well-powered online experiments ($N_1 = 437$, $N_2 = 310$, $N_3 = 166$), in which we systematically investigated effects of combining strategy instruction and feedback on students' learning. To test the generalizability of our findings, we used different feedback formats (Experiment 1: corrective feedback, Experiment 2: elaborated feedback), and experimentally varied the sequence of strategy instruction and feedback to rule out potential sequence effect of their combinations (Experiment 3). Additionally, in cases of reducing or non-additive effects (i.e., an instruction-by-feedback

interaction), we investigated the underlying processes of combining strategy instruction and feedback both from a cognitive, a metacognitive and an affective-motivational perspective (Narciss, 2008).

4.2 Strategy Instruction

Strategy instruction is regarded as an effective means to foster cognitive skill acquisition (Fyfe & Rittle-Johnson, 2016; Goldman, 1989). During strategy instruction the students are explicitly taught the subject matter on how to apply distinct strategies to support problem solving. To this end, strategy instruction often encompasses the use of worked examples (e.g., Fyfe & Rittle-Johnson, 2016; Goldman, 1989; Kant et al., 2017; Salden et al., 2010). Students first receive a short conceptual introduction that provides the basic conceptual knowledge referring to the underlying principles and concepts of the problem (e.g., Ohm's law in the domain of physics; Renkl, 2005; Wittwer & Renkl, 2008). Afterwards, students are provided with worked examples portraying the different solution steps leading to the final solution whereby their design follows principles of example selection and sequencing (Atkinson et al., 2000; Kant et al., 2017; Renkl, 2014; van Gog & Rummel, 2010).

From a cognitive load perspective, providing worked examples should reduce extraneous processing so that cognitive capacity is freed for germane cognitive processes, such as conceptual learning and elaboration (Kant et al., 2017; Renkl, 2014; van Gog & Paas, 2008). The effectiveness of instruction based on example-based learning is well-documented in several empirical studies (van Gog & Paas, 2008; see also Crissman, 2006; Hoogerheide & Roelle, 2020; Kirschner et al., 2006; Magliaro et al., 2005; Renkl, 1997, 2014, for empirical evidence). A recent meta-analysis in the domain of mathematics (Barbieri et al., 2023) also documented the benefits of worked examples compared to no-examples ($g = 0.48$), based on 181 effect sizes.

4.3 Feedback

In subsequent phases of skill acquisition, such as during practice phases, formative feedback is regarded as further fundamental instructional supplement to deepen students' learning (Graham et al., 2011, 2015; Hattie & Timperley, 2007; Narciss, 2012). Providing learners with feedback has shown to be effective for enhancing their performance (Hattie & Timperley, 2007) in both classrooms (e.g., Bangert-Drowns et al., 1991) and computer-based learning environments (e.g., Mertens et al., 2022; van der Kleij et al., 2015). Feedback can be considered as post-response instructional information about distinct aspects of a student's current performance or learning process to help the student to reduce discrepancies between the actual performance and the targeted performance (Hattie & Timperley, 2007; Narciss, 2008). It can be used to assess (e.g., in form of grades or scales), as well as to diagnose (e.g., whether a learning goal has been achieved), or to evaluate (e.g., the level of students' knowledge) performance regarding various success criteria (Hattie & Timperley, 2007; Kluger & DeNisi, 1996). The most important purpose of formative feedback is to guide the student to regulate and improve the further learning process (Hattie & Timperley, 2007; Narciss, 2008, 2012; Shute, 2008).

Regarding the feedback information provided by an external source (e.g., computer-based system), previous research investigated different types of feedback attributing different functions of feedback. Narciss (2012) differentiated between outcome-related feedback types, such as knowledge of result (KR), knowledge of correct response (KCR), or answer-until-correct (AUC) feedback (see also corrective feedback, Fyfe & Rittle-Johnson, 2016; Shute, 2008; lower-order feedback, Kuklick & Lindner, 2023; Mertens et al., 2022). These feedback types are related to technical characteristics of feedback to generate more correct responses. In order to improve incorrect responses, further information is often needed. Therefore, more elaborated feedback (EF) types are used that provide more complex feedback contents as compared to corrective or knowledge of result feedback (Kuklick &

Lindner, 2023; Mertens et al., 2022; Narciss, 2012; Shute, 2008). The use of elaborated feedback, which in addition to information on the correctness of the answer provides specific guidance on how to close existing knowledge gaps, is regarded to further enhance the effectiveness of feedback (Brooks et al., 2019; Hattie & Timperley, 2007; Mertens et al., 2022; Shute, 2008).

For instance, in their meta-analysis (435 studies, $k = 994$, $N > 61,000$ students), Wisniewski et al. (2020) investigated whether the type of feedback (reinforcement / punishment, corrective feedback, or elaborated feedback) moderated the effect of feedback on students' learning. Overall, the authors found a medium effect ($d = 0.48$) of providing feedback on learning. This effect was more pronounced for elaborated feedback ($d = 0.99$, $k = 42$ studies) than for corrective feedback ($d = 0.46$, $k = 238$ studies), or reinforcement / punishment ($d = 0.24$, $k = 39$ studies). Relatedly, Mertens et al. (2022) documented a comparable pattern for computer-based feedback. Based on 163 effects from 77 experimental studies they investigated by conducting a network meta-analysis whether different feedback types (KR, KCR, AUC, and EF compared with a no-feedback control condition) had different effects on lower- (i.e., recall/recognition) and higher-order (i.e., transfer) learning performance. Their results show that elaborated feedback is most effective to support learning (on lower-order: $g = 0.71$, $p < .01$, as well as on higher-order: $g = 0.46$, $p < .01$). Correct response feedback turned out to be least effective (only on lower-order: $g = 0.24$, $p = .026$; higher order: $g = 0.14$, $p = .345$).

4.4 Combining Strategy Instruction and Formative Feedback

Given that strategy instruction and formative feedback separately are both effective to support students' learning, they are often combined in educational practice (Magliaro et al., 2005; Rosenshine, 2008): Provided with strategy instruction, students should build a consistent mental representation of the problem situation and the corresponding solution

strategies. Receiving feedback during practice phases should further support the application and transfer of the previously learned solution strategies. Alternatively, the interactive, two-feedback-loop (ITFL) model by Narciss (2008) describes the implementation of feedback as a reciprocal interplay between internal (learner) and external (feedback) factors (cf. instructional medium) that may affect the effectiveness of feedback (see also Butler & Winne, 1995; Hattie & Timperley, 2007; Hayes, 2012; Panadero & Lipnevich, 2022; Winne & Butler, 1994, for related assumptions). Thus, alternatively, if prior strategy instruction (external factor) is given in advance of a practice phase, students could be inclined to undervalue the particular task requirements in the practice phase (see also illusion of understanding, Renkl, 2014). These subjective representations of task requirements could result in lower levels of processing of the feedback, both on the cognitive, metacognitive, or motivational level (Narciss, 2008). Thus, the combination could result in worse performance compared to the single interventions. Whether the combination of strategy instruction and feedback contributes more or even less to learning than a single intervention is still an open question.

Related evidence can be found in the synthesis by Salden et al. (2010). The authors synthesized the effects of eight studies in which strategy instruction (i.e., worked examples) and tutored problem solving comprising different feedback elements (e.g., prompts, immediate feedback, and context-sensitive hints) were combined. Across those studies, the findings indicated that adding strategy instruction to tutored problem solving contributed to learning.

However, there are also contradictory findings. To explore the effects of differences of prior strategy instruction and feedback, Fyfe and Rittle-Johnson (2016) provided elementary students with strategy instruction about solving mathematical equivalence problems or with control instruction. Afterwards, in a practice phase, students were asked to solve 12 structurally similar mathematical problems. One half of the elementary students received

corrective feedback after each mathematical problem whereas the other half did not receive feedback. The authors found a strategy-by-feedback interaction, as students only profited from strategy instruction when no feedback was provided (see also Fyfe & Rittle-Johnson, 2017, for similar findings). In Experiment 2, Fyfe and Rittle-Johnson (2016) provided all students with strategy instruction. Additionally, they varied whether students received no feedback, immediate feedback, or summative feedback on the practice tasks. Again, no feedback was more beneficial than providing feedback (see Shirah & Sidney, 2023, for related evidence; see also Wischgoll, 2017, for related findings in academic writing).

Together, these findings provided initial evidence that combining strategy instruction and formative feedback may not necessarily contribute to learning. Instead, they suggest that combining strategy instruction and feedback can also reduce learning performance or at least it has no additive effects (see also Kluger & DeNisi, 1996; Narciss, 2008; Shirah & Sidney, 2023).

4.4.1 Underlying Processes when Combining Strategy Instruction and Feedback

In line with the ITFL-model (Narciss, 2008), Fyfe and Rittle-Johnson (2016) proposed three non-mutually exclusive explanations for potential strategy instruction-by-feedback interactions.

Regarding the cognitive factors, after receiving strategy instruction, learners' subjective representation of the task requirements and in consequence, learners' processing of feedback might change due to increases of cognitive load, as students have to process redundant information of the strategy instruction and the feedback (Sweller et al., 2011).

Alternatively, regarding metacognitive factors, the provision of well-designed strategy instruction could result in an overestimation of students' comprehension having potentially down-side effects on the regulation of learning processes during subsequent practice phases as well as on their performance (see Baars et al., 2013; Finn & Metcalfe, 2014).

Last, regarding affective-motivational factors, the illusion of “feeling competent” may contribute to a lower level of arousal, and as such may be not beneficial to learning (Eccles & Wigfield, 2002; Grundmann et al., 2021; Kuklick & Lindner, 2023; Raaijmakers et al., 2017). Arousal refers to a state of excitement varying from low (unexcited, relaxed, calm) to high arousal (increased activation of the sympathetic nervous system; Arnsten, 2009; Hoogerheide et al., 2019). Previous research showed that the arousal level depends on the difficulty of the task or the valence of received feedback, that is, positive (i.e., the task was solved correctly) or negative (i.e., students’ answer was incorrect) feedback (Efklides & Dina, 2004; Kuklick & Lindner, 2023). Thus, the affective-motivational perspective may particularly account in cases of negative feedback, as negative feedback may disengage students (through failed improvements) and reduce learning performance (Grundmann et al., 2021).

However, it has to be noted that the studies that found a reducing effect of combining strategy instruction and feedback, predominantly applied lower-order feedback (Fyfe & Rittle-Johnson, 2016; Wischgoll, 2017). Thus, it is an open question whether these findings may generalize to more elaborated types of feedback. Therefore, the instruction-by-feedback interaction could be simply a result of the combination of strategy instruction and lower-order feedback, as lower-order feedback may not pose additional benefits regarding learning gains, and thus result in extraneous processing (see Mitchell & Jolley, 2010, for a critical discussion).

4.4.2 Sequence of Learning Phase and Practice Phase

Another aspect of the orchestration which might influence the combination of instruction and feedback is the sequence in which these activities are arranged. Recent research on PS-I (i.e., problem solving prior to instruction, see Kapur, 2012; and Loibl et al., 2017; Sinha & Kapur, 2021, for an overview) argued that preponing practice phases, such as problem

solving activities could enhance learning (see Sinha & Kapur, 2021, for meta-analytical evidence). Besides other mediating factors, the ability to notice potential flaws in students' understanding in an early stage of knowledge acquisition (Loibl & Rummel, 2014), is discussed as one underlying process that could explain the effects of flipping practice phases and instruction. Following Narciss (2008), flipping instruction and practice phases could furthermore enhance students' subjective representation of the task requirements, and thus result in more elaborated processing of the feedback.

Previous research regarding the effectiveness of PS-I, however, produced mixed findings (Kant et al., 2017; Loibl & Rummel, 2014; Sinha & Kapur, 2021; van Gog et al., 2011; van Harsel et al., 2020). On the one hand, there is supporting evidence for flipping instruction and practice phases: For instance, Sinha and Kapur (2021) conducted a meta-analysis of 53 studies including 166 effect sizes investigating the differential impact of the sequences problem solving followed by instruction (PS-I) versus instruction followed by problem solving (I-PS) on students' procedural, conceptual knowledge, and transfer performance in different domains. Their results showed a significant small to medium effect of PS-I ($g = .36$, 95% CI [0.20, 0.51]) regarding conceptual knowledge and transfer. Regarding procedural knowledge, in line with previous research, the authors found no significant effect ($g = -.03$, 95% CI [-0.20, 0.15]).

On the other hand, there is contrasting evidence in the domain of worked examples. For instance, van Gog et al. (2011) compared different sequences of instruction and practice phases: The authors presented secondary students ($N = 103$) either instruction-only (i.e., by providing worked examples), practice phase-only, instruction followed by a practice phase, or a practice phase followed by instruction. The results showed that the instruction-only condition and the instruction followed by a practice phase condition outperformed the practice phase-only condition and the practice phase followed by instruction condition (see also Kant et al., 2017, for related findings). However, whether these findings could be

transferred to practice phases incorporating feedback is an open question. It can be assumed that changing the sequence of instruction and feedback may contribute to learning. If students are engaged in a practice phase at the beginning of their studying, they could be more reliant on the additional feedback information to correctly solve the problems in the practice phase. The subsequent instruction phase would then serve as additional scaffold to consolidate the previously acquired strategies during the practice phase (see productive failure approach, Kapur, 2008).

4.5 Overview of the Present Study

The main aim of the present study, that included three online experiments, was to investigate potential (additive or reduced) effects of combining strategy instruction and feedback on students' physics learning in online environments. Throughout the experiments, we used identical learning materials (i.e., strategy instruction, practice tasks with feedback) to be able to compare findings across our experiments. In a multimedia learning environment, the students were provided with strategy instruction based on worked examples for solving electric circuits problems (Ohm's law, see Hoogerheide et al., 2019). In the practice phase, the students answered several practice tasks, in which they applied Ohm's law (by calculating the electric current) to different electric circuits.

The goal of Experiment 1 was to replicate a potential instruction-by-feedback interaction effect (as Fyfe & Rittle-Johnson, 2016, found in the domain of mathematics) in the domain of physics with corrective feedback. We used a 2×2-factorial design, crossing the two factors strategy instruction (yes vs. no) and feedback (yes vs. no). Furthermore, based on the ITFL-model by Narciss (2008), we tested for potential explanations regarding the underlying cognitive (i.e., mental effort, subjective difficulty), metacognitive (i.e., monitoring accuracy) and affective-motivational processes (i.e., arousal) accounting for additive versus reducing effects of combining strategy instruction and feedback.

Experiment 1

Given that instruction-by-feedback interaction could depend on the elaboration of feedback (external factor; Narciss, 2008), Experiment 2 aimed to generalize the findings and examine the robustness of the obtained effects in Experiment 1 by using elaborated feedback. In Experiment 3, we varied the sequence in which strategy instruction and feedback were provided, that is a learning phase with strategy instruction followed by a practice phase including feedback versus a preponed practice phase including feedback followed by a learning phase with strategy instruction. This variation allowed us to investigate whether instruction-by-feedback interaction may be a result of different sequences of strategy instruction and feedback. Together, these experiments allow to rigorously examine the underlying mechanisms and generalizability of combination effects of strategy instruction and feedback.

4.6 Experiment 1

4.6.1 Research Questions and Hypotheses

We pre-registered our hypotheses on as.predicted (see <https://aspredicted.org/tn6nx.pdf>). In line with previous evidence (Fyfe & Rittle-Johnson, 2016; Wischgoll, 2017), we hypothesized a main effect of strategy instruction on students' learning (Hypothesis 1a), as strategy instruction should help students construct a solid mental representation of the subject matter which should be conducive to learning. Regarding the effectiveness of corrective feedback (Experiment 1), it was less clear what to expect. Based on previous research, we can assume that different types of feedback should be differently effective (see Experiment 2 for details). However, in previous studies (e.g., Fyfe & Rittle-Johnson, 2016) on combination effects no significant main effect of feedback was obtained (which is in contrast to the common literature on the effectiveness of feedback). Thus, we refrained from making a clear prediction regarding main effects of feedback on learning.

Additionally, based on Fyfe and Rittle-Johnson (2016), we hypothesized an instruction-by-feedback interaction (Hypothesis 1b). Accordingly, the addition of feedback should impair the effectiveness of strategy instruction. We also explored whether the findings on transfer performance were robust for the underlying learning outcomes (i.e., near and far transfer).

In case we find an instruction-by-feedback interaction effect, we aimed at exploring the underlying processes of the potential instruction-by-feedback interaction from a cognitive, a metacognitive, and an affective-motivational perspective (Narciss, 2008). Following a cognitive perspective, we explored whether increases in cognitive load during the practice phase may mediate the effect of instruction-by-feedback interaction on learning. We therefore used subjective cognitive load ratings (i.e., mental effort, subjective difficulty) as potential mediators. Concretely, we assumed that students who received both instruction and feedback will experience an increase in cognitive load during the practice phase due to the dual support compared to students who received only one type of support (strategy instruction-only or feedback-only; Hypothesis 2a).

Following a metacognitive perspective, we also investigated whether students' overestimations mediate this effect. We used monitoring accuracy ratings to calculate students' overestimations: Students who rated their prospective performance as being better than their actual performance (self-estimated test performance > actual test performance) were defined as overestimated judgements. Specifically, we hypothesized that students who first received strategy instruction followed by feedback will perform worse than students who only received feedback due to their overestimation (represented by positive values of monitoring accuracy; Hypothesis 2b).

Last, following an affective-motivational perspective, we assumed that arousal mediates the effect. We used subjective ratings of perceived arousal to test this assumption. Concretely, we hypothesized that students who first received strategy instruction followed

by feedback in the practice phase will show lower arousal levels, resulting in poorer performance compared to students who received only feedback without any prior strategy instruction (Hypothesis 2c).

4.6.2 Method

4.6.2.1. *Participants*

In total, 499 students from non-physics study programs of a German university participated in the online experiment (which was approved by the local research ethics committee). The students were recruited via a university mailing list. We excluded students who were enrolled in physics, science and technology, or engineering ($n = 29$), whose native language was not German (or whose language level in German was below C1; $n = 0$), or who were not enrolled in any study program ($n = 3$). In addition, we only included those participants who solved both attention checks (see section 2.2.3. Materials, for details) correctly to be more restrictive ($n = 21$ failed at least one of the attention checks). Thus, the final sample consisted of $N = 437$ students.

The mean age of the students was 22.77 years ($SD = 3.80$) and 65% were female. The students were on average in their 6.68 semester ($SD = 4.74$) and most of them were enrolled in humanities ($n = 168$). The final sample size of 437 students exceeded the required sample size of $N = 157$, which was based on an a priori power analysis for an ANCOVA with interactions, expecting a medium to large effect size Cohen's $\eta^2 = .08$ (based on Fyfe & Rittle-Johnson, 2016) and power $(1 - \beta) = 0.95$. The higher number of students in our obtained sample resulted from an unexpected interest in our study (499 university students within 5 hours).

4.6.2.2. *Design*

We conducted a 2×2-factorial design with strategy instruction (yes vs. no) and feedback (yes vs. no) as between-participants factors. Participants were randomly assigned to one of

four conditions (without strategy instruction and without feedback, $n = 107$; without strategy instruction but with feedback, $n = 111$; with strategy instruction but without feedback, $n = 112$; with instruction and with feedback, $n = 107$).

We measured students' performance in the practice tasks, as well as their performance in the transfer tasks as dependent variables. Following the previously mentioned three different (but not mutually exclusive) perspectives on the underlying processes of an instruction-by-feedback interaction, we deliberately assessed subjective measures of cognitive load (mental effort, subjective difficulty), monitoring accuracy, and affect (arousal, pleasure) as potential cognitive, metacognitive and affective-motivational mediators. Furthermore, we collected data on prior knowledge, academic self-concept, and intrinsic motivation to better describe our online sample.

4.6.2.3. Materials

The entire experiment was presented with the online survey tool SoSci Survey (<https://www.soscisurvey.de/>). Based on previous learning and test materials (Rhöneck, 1986; Urban-Woldron & Hopf, 2012; see also Hoogerheide et al., 2019; Ivanjek et al., 2021), all the materials were carefully adapted by two experienced subject matter experts in physics education.

Prior Knowledge. We used a pretest comprising eight multiple-choice items (e.g., “What do you know about the total current in a parallel circuit?”; Cronbach's α for Experiment 1–3: $\alpha_1 = 0.56$, $\alpha_2 = 0.59$, $\alpha_3 = 0.53$) where students were asked to select one correct solution per item out of 4 possible options, adapted from commonly used test instruments in physics education to assess students' conceptual understanding of electrical circuits (Rhöneck, 1986; Urban-Woldron & Hopf, 2012; see also Hoogerheide et al., 2019; Ivanjek et al., 2021, for related approaches). For each correct answer, students received one point, yielding a

Experiment 1

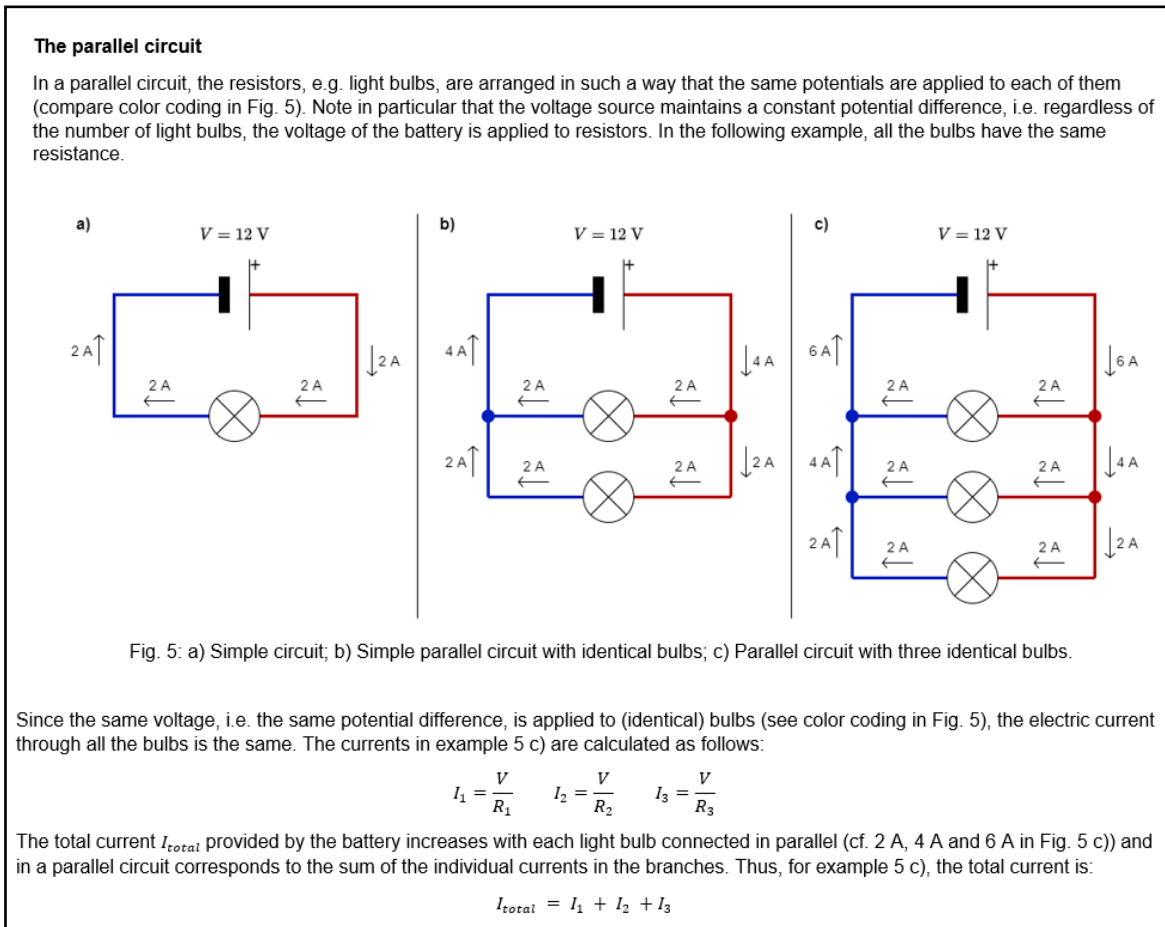
maximum score of eight points. We transformed the sum score to a proportion score (range: 0–1) by dividing the achieved points by the maximum points.

Strategy Instruction. In the learning phase, students were provided with either strategy instruction or no instruction. The main aim of the strategy instruction was to enable students to calculate currents in a parallel circuit on the basis of Ohm’s law in the form of $I = \frac{V}{R}$ with I referring to current, V to voltage, and R to resistance. Students needed to develop a conceptual understanding of this subject matter to solve the subsequent practice tasks.

The strategy instruction consisted of a multimedia presentation comprising conceptual information about the physical quantities “current”, “voltage”, and “resistance”, their relationship in electric circuits based on Ohm’s law as well as how a parallel circuit is constructed. A graphic representation of an electric circuit served to illustrate these abstract physics concepts. To enhance the effectiveness of the multimedia presentation, we considered general (multimedia) principles, such as signaling (Mautone & Mayer, 2001; Richter et al., 2018), and added process-oriented information to enhance conceptual understanding (van Gog et al., 2006, 2008; see also Lachner et al., 2019; see Figure 7a). In the worked example, the general solution steps were modeled to provide a strategy, how the current, displayed by the four ammeters in a parallel circuit (see Figure 7b), can be determined by applying Ohm’s law (see Hoogerheide et al., 2019).

Figure 7

Excerpt of the Strategy Instruction (Translated)

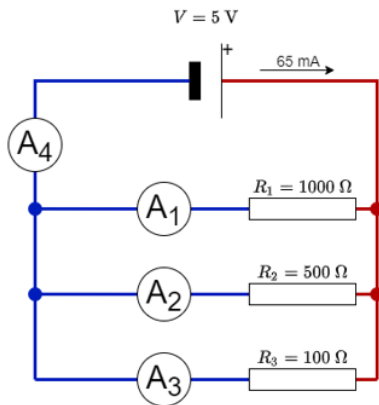


7a

Experiment 1

Example:

Read this example carefully.



Determine the current in ammeters A1 to A4 in the circuit shown above.

The same voltage is applied to all resistors of the parallel circuit, i.e. the same potential differences (see color coding). Depending on the respective resistors, these potential differences now cause an electric current in the branches according to:

$$I = \frac{V}{R}$$

$$I_1 = \frac{V}{R_1} = \frac{5\text{ V}}{1000\ \Omega} = 5\text{ mA} \quad I_2 = \frac{V}{R_2} = \frac{5\text{ V}}{500\ \Omega} = 10\text{ mA} \quad I_3 = \frac{V}{R_3} = \frac{5\text{ V}}{100\ \Omega} = 50\text{ mA}$$

The total current I_{total} measured in ammeter A4 results from the sum of the currents in the branches (I_1 , I_2 , I_3) corresponding to:

$$I_{total} = I_1 + I_2 + I_3$$

We therefore get:

$$I_{total} = 5\text{ mA} + 10\text{ mA} + 50\text{ mA} = 65\text{ mA}$$

We thus get the following currents through the ammeters:

A1 = 5 mA	A2 = 10 mA	A3 = 50 mA	A4 = 65 mA
-----------	------------	------------	------------

7b

Note. 7a shows the excerpt of the strategy instruction about how a parallel circuit is constructed and how the currents in the parallel circuit can be calculated using Ohm's law.

7b shows the worked example applying this knowledge.

In the no-strategy instruction conditions students received a control group instruction (see Figure 8; see Fyfe & Rittle-Johnson, 2016, for similar approaches). We included a control instruction instead of comparing the strategy instruction to a plain control condition which did not receive anything, to avoid potential differences regarding time-on-task (Möckel et

al., 2015). Differences in time-on-task may affect the overall amount of processing during the cognitive tasks, and potential differences in learning (Boksem et al., 2005, 2006; Langner et al., 2010; Lorist et al., 2000). Thus, holding time-on-task constant across condition can be regarded necessary to avoid potential confounds (Kovanović et al., 2015; Möckel et al., 2015).

Furthermore, we included a control instruction instead of another active treatment to be able to reach conclusion about the effectiveness of the strategy instruction. When different aspects of the instructions vary, we could not say with certainty whether the strategy instruction is responsible for the effect (Renkl, 2014). That is why it is important to keep the control instruction as constant as possible to the strategy instruction. To control for time-on-task for all conditions and nevertheless to avoid producing another treatment effect, we created a control instruction which would not influence the effect on task performance regarding the content (since the required information to solve the tasks was not included) but kept all other factors as constant as possible to the strategy instruction.

The control group instruction was comparable to the strategy instruction regarding text structure, schematics to illustrate the described contents, worked example, and terminology. The control group instruction also comprised a multimedia presentation (see Figure 8a) and a worked example (see Figure 8b). The contents of the strategy instruction and the control group instruction were divergent as students who received the control instruction were not introduced to Ohm's law or to the calculation of currents in a parallel circuit. Instead, the control instruction provided information about the invention of the battery and the effects of electric current. How to calculate currents in a parallel circuit was not explained in the multimedia presentation or in the worked example. Thus, the control instruction was not target-oriented for the following practice tasks but was only provided due to time-on-task constraints (see also Fyfe & Rittle-Johnson, 2016).

Experiment 1

Figure 8

Excerpt of the Control Instruction for the No-Instruction Condition (Translated)

Effects of the electric current

You cannot see an electric current. But there are signs that an electric current is flowing. This is because it can have different effects.

Probably the most obvious effect that can be observed is the **light effect**. When we switch on a light bulb, electricity flows through the light bulb and emits light. In other words: The light bulb becomes bright because of the current. In the process, the energy of the electric current is converted into light.

In old incandescent light bulbs, the current flows through a very thin wire, which subsequently heats up strongly. The glow of the wire causes light to be emitted. But this **thermal effect** of the electric current can also be used in other ways. For example, wires in irons or toasters are explicitly made to glow in order to use the generated heat.

In 1820, the physicist Hans Christian Oersted discovered the **magnetic effect** of electricity. In an experiment, he noticed that conductors with current flowing through them deflected magnetic needles in their vicinity. A magnetic field forms around a conductor in which a current flows. The magnetic field lines are arranged in concentric circles perpendicular to the conductor. This effect is used, for example, to lift heavy pieces of iron with an electromagnet.

Here, the right-fist rule applies: Grasp a conductor with your right fist so that the thumb points in the direction of the electric current (from + to -). Then the curved fingers indicate the direction of the magnetic field lines.

Another effect of electric current is that it can cause a **chemical reaction**. For example, electric current can be used to make water split into hydrogen (H₂) and oxygen (O₂). In the car industry, body parts are electrified during production to coat them with zinc, for example, to protect them from rust.

Note: Electric current can be recognized by its effects. There is the light effect, the heat effect, the magnetic effect and the chemical effect.

8a

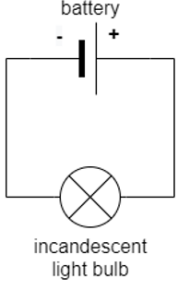
Example:

What effects of the electric current can be seen in the circuit on the right?

Solution:

The magnetic effect of the electric current can be seen in the leads from the positive terminal of the battery to the incandescent light bulb and back to the negative terminal of the battery.

In the light bulb, both the heat effect of the electric current and, as a result of the strong heating, the light effect of the current occurs.



battery

incandescent light bulb

8b

Note. 8a shows the excerpt of the control instruction about effects of electric current. 8b shows the worked example illustrating this knowledge.

Materials in the Practice Phase. The practice phase comprised two isomorphic practice tasks and additionally corresponding feedback in the feedback conditions. We have chosen this design because first, previous research showed that a restricted number of practice tasks were sufficient to foster students' learning. For example, Hoogerheide et al. (2019), whose material we adapted, only implemented one practice task. This decision is in line with findings by van Harsel et al. (2020), who showed that more (i.e., eight) tasks do not necessarily lead to more learning than fewer (i.e., four) tasks. Additionally, the use of a

restricted amount of closed-ended questions is often used for rapid assessments in online learning environments (cf. Kalyuga, 2006; Kalyuga & Sweller, 2004). Second, we aimed at reducing the drop-out risk (cf. attrition bias, Nunan et al., 2018) caused by the online study setting and the length of the questionnaire (Galesic, 2006; Hoerger, 2010). Thus, we decided to realize an effective and parsimonious implementation of guided practice and used two practice problems. Students answered these two isomorphic (i.e., analogous problem solving strategy, different surface features) practice tasks (Cronbach's α for Experiment 1–3: $\alpha_1 = 0.47$, $\alpha_2 = 0.56$, $\alpha_3 = 0.56$) in which they had to determine the currents by applying Ohm's law. Therefore, students had to select the correct solution from four alternatives. The distractors were based on common misconceptions students hold regarding currents (see McDermott & Shaffer, 1992; Schecker et al., 2018, for examples). The performance on the practice tasks was again transformed to proportion scores (range: 0–1).

We have chosen the multiple-choice test format (i.e., for the practice and transfer tasks) for several reasons: First, previous research showed that multiple-choice questions are at least as effective as other formats in consolidation phases (e.g., short-answer test; Adesope et al., 2017; Yang et al., 2021) as well as in combination with prior instruction (Ozuru et al., 2007).

Second, multiple-choice questions are frequently used in online learning environments combined with feedback because multiple-choice questions provide an easy as well as time- and resources-saving opportunity to analyze students' answers objectively and automatically by a computer-based system (Epstein et al., 2002; Guo et al., 2014; Kuechler & Simkin, 2003; Shin et al., 2019). That allows to provide immediate formative feedback to all students simultaneously which can increase students' self-regulation, learning gain, and motivation (Allen et al., 2016; Epstein et al., 2002; Guo et al., 2014; Shute, 2008). Thus, we used this advantage and provided students in the feedback conditions with immediate, corrective feedback by giving them information about whether their chosen answer was correct or not

Experiment 1

(Hattie & Timperley, 2007). Additionally, they received a brief justification why the selected answer was correct (see Figure 9a for the positive feedback) or why the chosen answer was incorrect (see Figure 9b for the negative feedback; KCR; Narciss, 2012). Through this short justification, students had the opportunity to expand or to consolidate their knowledge and to benefit from the feedback, even if they had chosen the correct answer.

Figure 9

Example of Feedback for Practice Task 1 (Translated)

Determine the amperage by ammeters A1 to A4 in the circuit shown above.

$I_1 = I_2 = 20 \text{ mA}$ and $I_3 = 40 \text{ mA}$ and $I_4 = 80 \text{ mA}$

$I_1 = I_2 = I_3 = I_4 = 80 \text{ mA}$

$I_1 = 40 \text{ mA}$ and $I_2 = 20 \text{ mA}$ and $I_3 = 20 \text{ mA}$ and $I_4 = 0 \text{ mA}$

$I_1 = 10 \text{ mA}$ and $I_2 = 5 \text{ mA}$ and $I_3 = 1 \text{ mA}$ and $I_4 = 16 \text{ mA}$

Your answer is correct! The voltage $V = 20 \text{ V}$ of the battery is applied to all resistors, but the resistance values differ. The currents through the individual resistors are given by $I = V/R$. The total current in A_4 is the sum of the currents through the individual resistors.

9a

Determine the amperage by ammeters A1 to A4 in the circuit shown above.

$I_1 = I_2 = 20 \text{ mA}$ and $I_3 = 40 \text{ mA}$ and $I_4 = 80 \text{ mA}$

$I_1 = I_2 = I_3 = I_4 = 80 \text{ mA}$

$I_1 = 40 \text{ mA}$ and $I_2 = 20 \text{ mA}$ and $I_3 = 20 \text{ mA}$ and $I_4 = 0 \text{ mA}$

$I_1 = 10 \text{ mA}$ and $I_2 = 5 \text{ mA}$ and $I_3 = 1 \text{ mA}$ and $I_4 = 16 \text{ mA}$

Your answer is wrong because the current is not the same in all parts of a parallel circuit! The voltage $V = 20 \text{ V}$ of the battery is applied to all resistors, but the resistance values differ. The currents through the individual are given by $I = V/R$. The total current in A_4 is the sum of the currents through the individual resistors. Thus, the correct answer would have been **$I_1 = I_2 = 20 \text{ mA}$ and $I_3 = 40 \text{ mA}$ and $I_4 = 80 \text{ mA}$.**

9b

Note. 9a shows the positive feedback students received if their chosen answer was correct.

9b shows the negative feedback students received if their chosen answer was incorrect.

Learning Outcomes (Posttest). Students' learning outcomes were assessed with four transfer tasks: two near transfer tasks and two far transfer tasks (McDonald's ω for overall transfer in Experiment 1–3: $\omega_1 = 0.57$, $\omega_2 = 0.56$, $\omega_3 = 0.49$). We adapted the transfer

instruments by Hoogerheide et al. (2019). To this end, the near transfer tasks were isomorphic to the problems from the practice tasks (i.e., “Determine the amperage by ammeters A1 to A4 in the electric circuit shown above”; Cronbach’s α for Experiment 1–3: $\alpha_1 = 0.60$, $\alpha_2 = 0.57$, $\alpha_3 = 0.67$) but based on different variable values. In the second task the circuit diagram was rotated. In the far transfer tasks, the students had to transfer their knowledge to different circuits with a different number of ammeters or different resistors: The first far transfer task was similar to the second near transfer task (rotated diagram) but with an additional ammeter. The second far transfer task seemed like the practice tasks and the first near transfer task with four ammeters but with a variable resistor. Since these two transfer items covered different facets of far transfer (additional ammeter, different type of resistance), it can be expected that the reliability would be insufficient (Cronbach’s α for Experiment 1–3: $\alpha_1 < 0.20$). Therefore, we conducted multilevel analyses as the items were nested within person (see section 4.6.2 Results).

Similar to the practice tasks, for each transfer task, students had to select the correct solution from four alternatives. The distractors were based on common misconceptions students hold regarding currents (see McDermott & Shaffer, 1992; Schecker et al., 2018, for examples). Students received one point per correct answer, resulting in a total score of 4 points for the transfer tasks. Again, we transformed the total score to a proportion score (range: 0–1). To explore, whether our findings were robust among knowledge facets, we also provide separate explorative analyses for the near and far transfer test components.

Perceived Cognitive Load. Students’ perceived cognitive load was measured by subjective ratings of students’ invested mental effort (“How much mental effort did you invest in studying the instruction/completing the task?”; Jacob et al., 2020; Paas, 1992), and their subjectively perceived task difficulty (“How difficult was it for you to understand the instruction/solve the task?”; DeLeeuw & Mayer, 2008; Jacob et al., 2020) after the learning

phase and after the practice phase on a Likert-scale from 1 (*not difficult at all*) to 9 (*very difficult*).

Monitoring Accuracy. To investigate potential differences of students' monitoring accuracy, students judged their expected performance on tasks about electric circuits ("How confident are you that you can solve tasks on the topic of electric circuits correctly?"; Eitel, 2016; Jacob et al., 2020; Nelson et al., 1994) on a scale from 0% (*not at all confident*) to 100% (*very confident*). Students rated their judgement of learning 3 times (i.e., before the pretest, after the learning phase, and after the practice phase). As in previous research, we operationalized monitoring accuracy as the difference between students' actual transfer performance and their prospective judgements of learning they have delivered after the learning phase (Eitel, 2016; Finn & Metcalfe, 2014). An overestimation of students' judged performance occurred when the values were positive, an underestimation occurred when the values were negative. A value of zero indicated correct assessment.

Affective Measures. We used the affective slider by Betella and Verschure (2016) to assess students' perceived arousal on a continuous scale from 1 (*sleepy/bored*) to 9 (*wide awake/concentrated*) and their perceived pleasure on a continuous scale from 1 (*sad*) to 9 (*happy*) as potential proxies for students' current affective states³ while learning.

Additional Control Measures.

Academic Self-Concept. As students' academic self-concept could impact their performance, we assessed their self-concept in the domain of physics by using five items (e.g., "I am good at physics.", "Physics is easy for me.", adapted from Gaspard et al., 2015, 2018; McDonald's $\omega = .93$) on a 4-point Likert-scale from 1 (*does not apply at all*) to 4 (*fully applies*).

³ We also measured perceived usefulness of the feedback during the practice phase as another indicator of students' motivational states (Eccles & Wigfield, 2002). However, due to technical issues, the scale was only visible for two of the four conditions. We, therefore, did not consider it in Experiment 1.

Intrinsic Motivation. We measured students' intrinsic motivation and interest in physics with five items (e.g., "I enjoy physics.", "I am interested in physical topics.", adapted from Gaspard et al., 2015, 2018; based on Pekrun et al., 2002; McDonald's $\omega = .89$) on a 4-point Likert scale from 1 (*does not apply at all*) to 4 (*fully applies*).

Attention Checks. In our experiment, we included an attention check before and after the practice phase to control for students' attention while completing the online experiment and increase quality of the obtained data and implementation fidelity (see also Hauser & Schwarz, 2016; Oppenheimer et al., 2009, for related approaches). Students were asked to fill in two missing numbers in a very simple 4×4 Sudoku. Instructions about how to solve a Sudoku were provided beforehand.

4.6.2.4. Procedure

At the beginning, we informed the students about the aim of the experiment. After providing their written consent, students answered the demographic questionnaire (i.e., demographics, academic self-concept, intrinsic motivation, current arousal, pleasure, and their first judgement of learning). Afterwards, all students answered the pretest. Then, the students were randomly assigned to one of the four conditions (without strategy instruction and without feedback, without strategy instruction but with feedback, with strategy instruction but without feedback, with both strategy instruction and feedback). In the learning phase, students randomly read one of the two versions of the multimedia instruction (strategy instruction vs. no instruction), and assessed their perceived mental effort, subjective difficulty, second judgement of learning, and their current arousal and pleasure. After the first attention check in the practice phase, students solved the two practice tasks, students in the feedback conditions additionally received feedback (see Figure 9). After the practice phase, all students again rated their perceived mental effort, subjective difficulty, third judgement of learning, current arousal, and current pleasure. After completing the

second attention check, all students solved the posttest. The study lasted about 30 minutes and was rewarded with 6 Euros (≈ 6 \$).

4.6.3 Results

As effect size measures, we used partial η_p^2 with qualifying values of $\eta_p^2 = .01, .06, .14$ as small, medium, and large effects (Cohen, 2013). Moreover, we used an alpha level of $\alpha = .05$. In cases of computing generalized mixed effect models, as estimation method, we used maximum likelihood (Laplace approximation), logit as the link function, and Akaike information criterion (AIC), Bayesian information criterion (BIC), and Log likelihood (LogLik) as goodness-of-fit method (cf. Bono et al., 2021).

4.6.3.1. Preliminary Analyses

Initial boxplot analyses indicated no extreme outliers. Correlations between prior knowledge, the dependent variables and the mediation variables are presented in Appendix A. Separate analyses of variance (ANOVAs) showed that conditions did not differ regarding academic self-concept, $F(3, 433) = 1.36, p = .253, \eta_p^2 = .01$, intrinsic motivation, $F < 1$, baseline arousal, $F(3, 433) = 1.75, p = .155, \eta_p^2 = .01$, baseline pleasure, $F(3, 433) = 1.60, p = .189, \eta_p^2 = .01$, baseline monitoring accuracy, $F(3, 433) = 1.64, p = .179, \eta_p^2 = .02$, prior knowledge, $F < 1$, and time-on-task, $F(3, 433) = 1.02, p = .385, \eta_p^2 = .01$. The descriptive values of all measurements can be seen in Table 8.

Table 8*Means and Standard Deviations for all Measurements in Experiment 1*

Variables	Without instruction, without feedback (<i>n</i> = 107)	Without instruction, with feedback (<i>n</i> = 111)	With instruction, without feedback (<i>n</i> = 112)	With instruction, with feedback (<i>n</i> = 107)
Pretest				
Prior knowledge (0–1)	0.60 (0.22)	0.59 (0.21)	0.59 (0.21)	0.58 (0.22)
Practice phase				
Practice tasks (0–1)	0.61 (0.39)	0.63 (0.38)	0.75 (0.37)	0.69 (0.37)
Posttest				
Transfer tasks (0–1)	0.56 (0.29)	0.61 (0.30)	0.62 (0.28)	0.65 (0.29)
Near transfer tasks (0–1)	0.74 (0.36)	0.74 (0.39)	0.71 (0.37)	0.81 (0.33)
Far transfer tasks (0–1)	0.38 (0.32)	0.47 (0.34)	0.53 (0.33)	0.50 (0.36)
Mediator variables				
Mental effort (after instruction) (1–9)	5.36 (1.69)	5.50 (1.95)	5.70 (1.84)	5.61 (2.06)
Mental effort (after practice phase) (1–9)	5.78 (1.95)	5.49 (2.19)	5.17 (1.88)	4.96 (2.01)
Subjective difficulty (after instruction) (1–9)	3.66 (2.17)	4.60 (2.39)	4.68 (2.44)	4.77 (2.52)
Subjective difficulty (after practice phase) (1–9)	6.27 (2.46)	5.37 (2.61)	4.93 (2.64)	4.28 (2.55)
Monitoring accuracy (baseline) (–100%–100%)	13.67 (24.96)	6.36 (24.43)	9.39 (24.86)	10.29 (23.77)
Monitoring accuracy (after instruction) (–100%–100%)	–15.83 (38.78)	–26.52 (39.42)	–36.40 (40.71)	–28.69 (36.99)
Monitoring accuracy (after practice phase) (–100%–100%)	–11.16 (30.41)	–24.50 (35.30)	–23.68 (34.74)	–24.95 (30.32)
Arousal (baseline) (1–9)	5.13 (1.80)	4.65 (1.84)	5.09 (1.71)	4.81 (1.91)
Arousal (after instruction) (1–9)	5.10 (1.64)	4.86 (1.70)	5.10 (1.90)	5.02 (1.86)
Arousal (after practice phase) (1–9)	5.15 (1.84)	5.14 (2.01)	5.09 (2.03)	5.21 (1.97)
Pleasure (baseline) (1–9)	6.05 (1.67)	5.57 (1.78)	5.77 (1.71)	5.64 (1.80)
Pleasure (after instruction) (1–9)	5.95 (1.66)	5.54 (1.70)	5.83 (1.65)	5.63 (1.64)
Pleasure (after practice phase) (1–9)	5.81 (1.67)	5.61 (1.74)	5.80 (1.75)	5.69 (1.71)

Experiment 1

Variables	Without instruction, without feedback (<i>n</i> = 107)	Without instruction, with feedback (<i>n</i> = 111)	With instruction, without feedback (<i>n</i> = 112)	With instruction, with feedback (<i>n</i> = 107)
Control variables				
Self-concept (1–4)	2.38 (0.76)	2.20 (0.81)	2.21 (0.70)	2.29 (0.69)
Intrinsic motivation (1–4)	2.55 (0.73)	2.50 (0.73)	2.45 (0.73)	2.48 (0.65)
Time in total (in seconds)	1214.01 (420.32)	1321.59 (526.60)	1273.79 (509.66)	1240.95 (453.03)

4.6.3.2. *Performance on the Practice Tasks*

We aimed to explore potential differences in students' performance and to examine potential main effects of strategy instruction or feedback, or an interaction effect of both interventions. We realized generalized mixed effect models applying binary logistic regressions with participants as random factor, instruction, feedback and the interaction of instruction and feedback as fixed factors and students' performance on the practice tasks as the dependent variable nested within participants (and repeated because of two scores on practice tasks). Additionally, we controlled for students' prior knowledge.

The multilevel analysis showed a significant main effect of strategy instruction, indicating that students who received strategy instruction solved the practice tasks better than students who did not receive the strategy instruction. There was no main effect of feedback and no significant interaction effect (see Table 9 for the entire test statistics)⁴.

⁴ As pre-registered, we also conducted separate ANCOVAs for each of the learning outcomes with strategy instruction (yes vs. no) \times feedback (yes vs. no) as between-participants factors, students' respective task performance as dependent variable and controlled for students' prior knowledge included as covariate. Since the reliability of some of the outcome variables were insufficient (Cronbach's $\alpha < .50$), we applied multilevel analysis as additional safeguard. To be succinct, we decided to only report the results of the multilevel analyses. If the ANCOVAs and the multilevel analyses produced different effects, we explicitly marked them in the body text. The entire findings of the ANCOVAs can be seen in Appendix B.

Experiment 1

Table 9

Results of the Multilevel Analysis Exploring the Effects of Feedback, Strategy Instruction, and its Interaction on Students' Practice Tasks, Overall, Near, and far Transfer Performance in Experiment 1

Variables	Estimate	SE _b	<i>p</i>
Practice tasks			
Strategy Instruction	.89	0.30	.003
Feedback	.12	0.28	.677
Interaction	-.45	0.41	.264
Transfer tasks (overall)			
Strategy Instruction	.68	0.22	.002
Feedback	.39	0.22	.071
Interaction	-.52	0.31	.089
Near transfer			
Strategy Instruction	-.05	0.32	.872
Feedback	.01	0.32	.974
Interaction	.31	0.46	.502
Far transfer			
Strategy Instruction	.73	0.21	.001
Feedback	.46	0.21	.033
Interaction	-.60	0.30	.044

Note. Significant *p*-values are highlighted in bold letters.

4.6.3.3. *Performance on the Transfer Tasks*

We applied the same multilevel analysis as regarding students' performance on the practice tasks but now with students' performance on the transfer tasks as dependent variable. In line with Hypothesis 1a, we obtained a significant main effect of strategy instruction. The main effect of feedback, as well as the instruction-by-feedback interaction effect were not significant (see Table 9).

4.6.3.4. *Additional Explorative Analyses*

To explore differences between near and far transfer, we additionally conducted explorative separate analyses for near and far transfer to investigate whether instruction and feedback had differential effects on near and far transfer. Therefore, we followed the same procedure but first with students' performance in the near transfer tasks as the dependent variable of the multilevel analysis and in a second separate multilevel analysis students' performance in the far transfer tasks as the dependent variable.

Regarding students near transfer performance as the outcome variable, the results of the multilevel analysis showed that there were no significant main effects neither for strategy instruction nor for feedback, as well as no significant interaction effect (see Table 9). We did not obtain significant differences among conditions, $F(3, 433) = 1.51, p = .212, \eta_p^2 = .01$.

In contrast to near transfer, regarding far transfer, we obtained a significant main effect of strategy instruction (see Table 9), indicating that students who received strategy instruction but no feedback ($M = 0.53, SD = 0.33$) outperformed students who received neither strategy instruction nor feedback ($M = 0.38, SD = 0.32, p = .005, d = 0.46$). Furthermore, we found a significant main effect of feedback (see Table 9), indicating that students who received feedback outperformed students who received no feedback. However, the main effect of feedback seemed to be not that robust, as the effect was not significant, when we performed an ANCOVA (see Appendix B).

Experiment 1

Regarding the interaction effect of strategy instruction and feedback, the multilevel analysis obtained a significant instruction-by-feedback interaction effect on students' far transfer performance (see Table 9).

To break up the significant interaction, we conducted simple effect analyses regarding far transfer. They revealed that when there was no feedback provided, the presence of strategy instruction significantly improved far transfer test performance ($p = .007$; receiving strategy instruction: $M = 0.53$, $SD = 0.33$; without strategy instruction: $M = 0.38$, $SD = 0.32$). By contrast, when participants received feedback, there was no effect of strategy instruction ($d = 0.09$, $p = .861$; receiving strategy instruction: $M = 0.50$, $SD = 0.36$; without strategy instruction: $M = 0.47$, $SD = 0.34$). Together, the findings suggest that strategy instruction had an effect on learning only when no additional feedback was provided. An instruction-by-feedback interaction only emerged for higher-order learning outcomes (i.e., far transfer). The results further suggested that additional corrective feedback had no additive effect on the far transfer performance.

To explore the underlying processes which could explain the significant instruction-by-feedback interaction effect on far transfer, based on our pre-registration and the ITFL-model (Narciss, 2008), we explored cognitive (i.e., mental effort, subjective difficulty), metacognitive (i.e., monitoring accuracy), and affective-motivational processes (i.e., arousal, pleasure) by applying separate simple mediation analyses with one of the mediator variables each. We represented the instruction-by-feedback interaction by creating a new variable named "interaction" which we included as contrast-coded predictor in the regression model (Rosnow & Rosenthal, 1996; Wiens & Nilsson, 2017). We assigned the following contrast weights of the interaction variable according to our theoretical assumptions: $-1 =$ without instruction without feedback, $1 =$ without instruction with feedback, $1 =$ with instruction without feedback, $-1 =$ with instruction with feedback (cf. Lachner et al., 2020). Students' far transfer performance was the dependent variable. We used the PROCESS

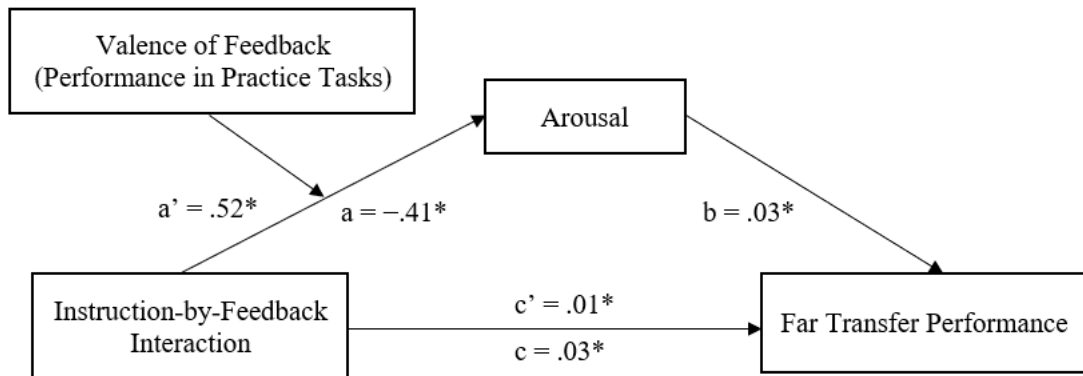
macro (Version 4) in SPSS (Hayes, 2018) with bootstrapped 95% confidence intervals (10,000 bootstrap samples). None of the univariate mediation analyses approached significance (Appendix C shows the results of all mediation analyses).

In a next step, we conducted additional moderated mediation analyses including valence as moderator variable, as indicated by students' performance on the practice tasks, following theoretical suggestions of Grundmann et al. (2021) that negative feedback in particular (i.e., feedback, when the answer was incorrect) could largely affect the proposed mediators (see also Eccles & Wigfield, 2002; Kalyuga & Sweller, 2014; Kluger & DeNisi, 1996; Raaijmakers et al., 2017) and reduce learning performance due to students' disengagement.

Based on these theoretical assumptions, we performed moderated mediation analyses in which a moderator (i.e., W = valence) moderates the a-path of the mediation model (independent variable to mediator) but not the c'-path (direct effect) based on Model 7 of Hayes' (2018) PROCESS macro for SPSS. Interestingly, we found a significant effect of arousal, $b = 0.01$, $SE_b = 0.00$ (95%, CI [0.0005, 0.0154]; see Figure 10), as zero was not included in the confidence interval. More precisely, valence moderated the effect of interaction on arousal, suggesting that depending on the experimental condition (instruction only, feedback only, both, or neither), students' valence differed. This affected student arousal and, subsequently, far transfer. None of the other moderated mediation analyses revealed significance (see Appendix D). In summary, the results indicated that the instruction-by-feedback interaction may result in far transfer tasks due to lower levels of arousal. This effect, however, only held for students who received negative feedback during the practice phase.

Figure 10

Moderated Mediation Analysis Regarding the Impact of the Valence of Feedback and Students' Arousal



Note. $*p < .05$.

4.6.4 Discussion

In Experiment 1, in line with Hypothesis 1a, we found a main effect of strategy instruction on students' learning outcome. Overall, we did not obtain a main effect of feedback or an instruction-by-feedback interaction effect on students' transfer performance. More fine-grained exploratory analyses in which we investigated students' near and far transfer performance revealed a significant instruction-by-feedback interaction for far transfer. The combination of strategy instruction and feedback did not contribute more to learning than the single interventions (cf. Shirah & Sidney, 2023, for related findings).

Further analyses of the underlying processes of the instruction-by-feedback interaction were rather in line with our affective-motivational assumptions. Students who received strategy instruction and negative feedback perceived lower levels of arousal during the feedback. Lower levels of arousal were associated with lower far transfer performance. This pattern, however, only emerged for students who received negative feedback in the practice phase.

The findings of Experiment 1 therefore suggest that instruction-by-feedback interactions only emerge under distinct conditions, namely when corrective feedback was used, only with regard to high-level learning outcomes (i.e., far transfer), and for students who received negative feedback, as they perceived lower levels of arousal. As in the study by Fyfe and Rittle-Johnson (2016), in Experiment 1, we realized rather parsimonious feedback, as students only received corrective feedback and a short-answer explanation, as to why their solution was (in-)correct.

The interaction effect can be explained by students' arousal as a consequence of negative feedback, as similar to recent research by Kuklick and Lindner (2023), the task performance (i.e., valence) moderated the emotional impact of feedback (i.e., arousal). Grundmann et al. (2021) also showed that negative feedback can lead to motivated disengagement, which hinders performance improvements. Thus, the arousal resulting from the negative feedback could have led to loss of motivation and feedback disengagement, which could be another explanation for the reducing effect.

Kuklick and Lindner (2023) found that after correct response feedback, students reported more negative emotions than when, for example, elaborated feedback was provided. Furthermore, their findings showed that elaborated feedback increased the perceived usefulness. It is unclear whether the effects of Experiment 1 would replicate (with more robust effects) in settings in which more elaborated feedback is provided. When students would receive more specific information in the feedback, the perceived utility of the feedback and the associated arousal could increase, which could result in higher levels of test performance (Shute, 2008), and a reversal into an additive combination effect.

4.7 Experiment 2

In Experiment 2, we aimed at generalizing the previous findings to more elaborated feedback (see section 4.7.1.2 Design and Material). We hypothesized a main effect for

instruction (Hypothesis 1a) and a main effect for elaborated feedback (Hypothesis 1b). Furthermore, we hypothesized an instruction-by-feedback interaction effect (Hypothesis 2).

4.7.1 Method

4.7.1.1. *Participants*

In total, 417 students from non-physics study programs of a German university participated in Experiment 2. We applied the identical exclusion criteria of Experiment 1 (students who were enrolled in physics, engineering, science and technology, or technics: $n = 23$; students who failed both attention checks: $n = 4$; students who were not enrolled in study programs: $n = 3$). In addition, to obtain robust estimates, we excluded students who had taken physics, technology and science, or technics as a major subject in high school to ensure that the sample consists of novice students ($n = 46$). We further excluded students who had already participated in the first experiment ($n = 18$). And we only included those students who solved both attention checks correctly ($n = 13$ failed at least one of the two attention checks). Thus, the final sample comprised $N = 310$ students as scheduled. The mean age of the students was 22.95 years ($SD = 3.63$) and 67% were female. The students were on average in their 7.19 semester ($SD = 4.84$) and most of them were enrolled in humanities in at least one subject ($n = 169$). The students were comparably distributed among conditions (without instruction without feedback, $n = 81$; without instruction with feedback, $n = 77$; with instruction without feedback, $n = 78$; with instruction with feedback, $n = 74$).

4.7.1.2. *Design and Material*

We implemented a 2×2 -factorial design with strategy instruction (yes vs. no) and feedback (yes vs. no) as between-participants factors. Participants were randomly assigned to one of four conditions (without strategy instruction and without feedback, $n = 81$; without strategy instruction but with feedback, $n = 77$; with strategy instruction but without feedback, $n = 78$; with instruction and with feedback, $n = 74$).

We used the same material as in Experiment 1 except for the feedback, which was provided also on the process level (Hattie & Timperley, 2007; Narciss, 2013). Students were provided with remedial explanations to additionally enhance their conceptual understanding: Thus, in the feedback the underlying physics principles (i.e., how voltage, resistances and currents are related in a parallel circuit and how to calculate the different partial currents on the basis of voltage and resistances) were explained. Additionally, we showed the correct calculations to enhance their procedural understanding (see Figure 11). Thus, the feedback comprised more specific information regarding about which solution steps to perform, and in what order, to solve the task correctly, and thus constitutes one possible form of elaborated feedback (Shute, 2008).

Figure 11

Example of Negative Feedback of Practice Task 1 in Experiment 2, if the Answer was Incorrect (Translated)

Determine the amperage by ammeters A1 to A4 in the circuit shown above.

- $I_1 = I_2 = 20 \text{ mA}$ and $I_3 = 40 \text{ mA}$ and $I_4 = 80 \text{ mA}$
- $I_1 = I_2 = I_3 = I_4 = 80 \text{ mA}$
- $I_1 = 40 \text{ mA}$ and $I_2 = 20 \text{ mA}$ and $I_3 = 20 \text{ mA}$ and $I_4 = 0 \text{ mA}$
- $I_1 = 10 \text{ mA}$ and $I_2 = 5 \text{ mA}$ and $I_3 = 1 \text{ mA}$ and $I_4 = 16 \text{ mA}$

Your answer is wrong! ✘ The voltage $V = 20 \text{ V}$ of the battery is applied to all resistors, but the resistance values differ. This means that the current in the individual resistors can differ. In the branch of A₁ and A₂, the resistance is 1000Ω in each case, so $I = V/R$ results in a current of $I_1 = I_2 = \frac{20 \text{ V}}{1000 \Omega} = 20 \text{ mA}$ in each case. At the resistor in the third branch with 500Ω , the corresponding current is $I_3 = \frac{20 \text{ V}}{500 \Omega} = 40 \text{ mA}$. The total current in A₄ results from the sum of the currents through the individual branches of the parallel circuit. $I_4 = I_1 + I_2 + I_3 = 20 \text{ mA} + 20 \text{ mA} + 40 \text{ mA} = 80 \text{ mA}$. Thus, the correct answer would have been **$I_1 = I_2 = 20 \text{ mA}$ and $I_3 = 40 \text{ mA}$ and $I_4 = 80 \text{ mA}$.**

Perceived Utility of the Practice Phase. Another modification was that all students indicated their perceived utility of the practice phase (inclusive feedback). To assess the utility of the practice phase, we used five items (e.g., “I have considered the practice phase to be helpful”; McDonald’s $\omega = .81$) based on Gaspard et al. (2015) on a scale from 1 (*not at all useful*) to 9 (*very useful*).

4.7.1.3. Procedure

The procedure was the same as in Experiment 1.

4.7.2 Results

4.7.2.1. Preliminary Analyses

Initial boxplot analyses indicated no extreme outliers. Correlations between prior knowledge, the dependent variables and the mediation variables are presented in Appendix E. As in Experiment 1, students were comparable among conditions regarding their academic self-concept, $F < 1$, intrinsic motivation, $F(3, 306) = 1.06, p = .364, \eta_p^2 = .01$, their baseline arousal, $F(3, 306) = 1.29, p = .279, \eta_p^2 = .01$, their baseline pleasure, $F(3, 306) = 2.27, p = .080, \eta_p^2 = .02$, their baseline monitoring accuracy, $F(3, 304) = 1.46, p = .226, \eta_p^2 = .01$, their prior knowledge, $F(3, 306) = 1.70, p = .167, \eta_p^2 = .02$, and regarding time-on-task, $F < 1$. The descriptive values of all measurements can be seen in Table 10.

Table 10*Means and Standard Deviations for all Measurements in Experiment 2*

Variables	Without instruction, without feedback (<i>n</i> = 81)	Without instruction, with feedback (<i>n</i> = 77)	With instruction, without feedback (<i>n</i> = 78)	With instruction, with feedback (<i>n</i> = 74)
Pretest				
Prior knowledge (0–1)	0.28 (0.12)	0.31 (0.10)	0.27 (0.14)	0.29 (0.12)
Practice phase				
Practice tasks (0–1)	0.60 (0.43)	0.69 (0.36)	0.72 (0.38)	0.74 (0.35)
Posttest				
Transfer tasks (0–1)	0.54 (0.32)	0.67 (0.25)	0.66 (0.30)	0.72 (0.28)
Near transfer tasks (0–1)	0.67 (0.39)	0.82 (0.32)	0.72 (0.37)	0.82 (0.31)
Far transfer tasks (0–1)	0.41 (0.33)	0.52 (0.31)	0.59 (0.36)	0.61 (0.36)
Mediator variables				
Mental effort (after instruction) (1–9)	5.68 (1.95)	5.56 (1.90)	5.82 (1.95)	5.58 (1.90)
Mental effort (after practice phase) (1–9)	6.23 (1.95)	5.73 (2.14)	5.64 (2.28)	4.77 (2.00)
Subjective difficulty (after instruction) (1–9)	3.79 (2.06)	4.08 (2.28)	5.60 (2.40)	4.81 (2.44)
Subjective difficulty (after practice phase) (1–9)	6.86 (2.26)	4.88 (2.49)	5.28 (2.47)	4.09 (2.47)
Monitoring accuracy (baseline) (–100%–100%)	7.39 (24.15)	4.52 (25.33)	9.00 (25.18)	6.02 (20.86)
Monitoring accuracy (after instruction) (–100%–100%)	–20.21 (43.34)	–26.01 (37.09)	–30.35 (38.45)	–27.77 (32.99)
Monitoring accuracy (after practice phase) (–100%–100%)	–26.26 (30.50)	–20.38 (31.53)	–21.29 (32.98)	–20.73 (28.69)
Arousal (baseline) (1–9)	5.26 (1.70)	4.84 (1.91)	4.99 (1.95)	4.70 (1.84)
Arousal (after instruction) (1–9)	5.26 (1.74)	5.36 (1.93)	5.06 (2.13)	5.18 (1.97)
Arousal (after practice phase) (1–9)	5.30 (1.91)	5.53 (2.09)	5.22 (2.14)	5.32 (1.92)
Pleasure (baseline) (1–9)	6.01 (1.61)	5.71 (1.92)	5.54 (1.73)	5.31 (1.66)
Pleasure (after instruction) (1–9)	5.84 (1.46)	5.69 (1.80)	5.51 (1.83)	5.41 (1.53)
Pleasure (after practice phase) (1–9)	5.40 (1.77)	5.86 (1.97)	5.62 (1.72)	5.66 (1.58)

Experiment 2

Variables	Without instruction, without feedback (<i>n</i> = 81)	Without instruction, with feedback (<i>n</i> = 77)	With instruction, without feedback (<i>n</i> = 78)	With instruction, with feedback (<i>n</i> = 74)
Control variables				
Utility of the learning phase (1–9)	6.37 (1.46)	6.38 (1.45)	6.56 (1.63)	6.50 (1.42)
Utility of the practice phase (1–9)	4.55 (1.93)	6.39 (1.61)	5.56 (1.67)	6.05 (1.79)
Self-concept (1–4)	2.15 (0.74)	2.24 (0.69)	2.11 (0.73)	2.13 (0.74)
Intrinsic motivation (1–4)	2.48 (0.65)	2.46 (0.64)	2.34 (0.75)	2.32 (0.71)
Time in total (in seconds)	1392.32 (513.07)	1317.91 (469.48)	1339.27 (601.89)	1290.73 (433.72)

4.7.2.2. Performance on the Practice Tasks

As in Experiment 1, we aimed to explore potential differences in students' performance and to examine potential main effects of strategy instruction or feedback, or an interaction effect of both interventions. We applied the same analyses as in Experiment 1. The multilevel analyses showed that none of the main and interaction effects were significant (see Table 11 for the entire test statistics)⁵.

⁵ The results of the multilevel analyses confirmed the results of the ANCOVAs. The results of the ANCOVAs can be seen in Appendix F.

Experiment 2

Table 11

Results of the Multilevel Analysis Exploring the Effects of Feedback, Strategy Instruction, and its Interaction on Students' Practice Tasks, Overall, Near, and far Transfer Performance in Experiment 2

Variables	Estimate	SE _b	<i>p</i>
Practice tasks			
Strategy Instruction	.83	.43	.055
Feedback	.01	.43	.984
Interaction	-.07	.61	.905
Transfer tasks (overall)			
Strategy Instruction	.54	.23	.017
Feedback	.47	.23	.039
Interaction	-.12	.33	.707
Near transfer			
Strategy Instruction	.15	.32	.641
Feedback	.98	.36	.007
Interaction	-.14	.51	.785
Far transfer			
Strategy Instruction	.73	.25	.004
Feedback	.37	.25	.143
Interaction	-.22	.35	.542

Note. Significant *p*-values are highlighted in bold letters.

4.7.2.3. Performance on the Transfer Tasks

We proceeded similarly for the performance on the transfer tasks (learning outcome). The multilevel analysis revealed a significant main effect of strategy instruction (see Table 11), indicating that students who received strategy instruction outperformed students who did not receive strategy instruction in the transfer tasks. Contrarily to Experiment 1, we found a significant main effect of elaborated feedback (see Table 11), indicating that students who received feedback during the practice phase performed better in the following transfer tasks than students who did not receive feedback. The instruction-by-interaction effect did not approach significance (see Table 11), indicating that the combination did not lead to detrimental or differential effects. Instead, the fact that strategy instruction as well as feedback had an effect, indicated that the combination of strategy instruction and feedback led to additive effects.

4.7.2.4. Additional Explorative Analyses

In line with Experiment 1, we conducted fine-grained explorative analyses to determine whether analyzing near and far transfer performances separately yielded different results. We followed the same procedure of analysis but first with students' performance in the near transfer tasks as the dependent variable of the multilevel analysis and in a second separate multilevel analysis students' performance in the far transfer tasks as the dependent variable.

Regarding students' near transfer performance, we found a significant main effect of feedback (see Table 11; without strategy instruction but with feedback: $M = 0.82$, $SD = 0.32$; without strategy instruction and without feedback: $M = 0.67$, $SD = 0.39$). However, there was no main effect of strategy instruction and no significant interaction effect (see Table 11).

In contrast to the results regarding the near transfer performance, regarding far transfer, results revealed a significant main effect of strategy instruction, (see Table 11; with strategy

Experiment 2

instruction but without feedback: $M = 0.59$, $SD = 0.36$; without strategy instruction and without feedback: $M = 0.41$, $SD = 0.33$). The effect of feedback and the instruction-by-feedback interaction effect were not significant (see Table 11). Apparently, instruction rather affected far transfer performance, whereas feedback affected near transfer performance. Contrarily to Experiment 1, the interaction effect on far transfer was no longer significant, likely because we provided more elaborated feedback.

As another exploratory analysis, to explore potential differences among conditions regarding students' perceived utility of the practice phase, we performed an ANOVA with strategy instruction and feedback as between-participants factors and students' perceived utility of the practice phase as dependent variable. Results showed a significant positive main effect for feedback, $F(1, 302) = 34.39$, $p < .001$, $\eta_p^2 = .10$, and a significant instruction-by-feedback interaction effect, $F(1, 302) = 11.22$, $p = .001$, $\eta_p^2 = .04$.

To break up the significant effects, we computed simple effect analyses. The analyses revealed that when students received no strategy instruction, they found the feedback ($M = 6.39$, $SD = 1.61$) in the practice phase more useful than no feedback ($M = 4.55$, $SD = 1.93$, $p < .001$). This effect disappeared for the strategy instruction conditions, as the practice phases with ($M = 6.05$, $SD = 1.79$) and without feedback ($M = 5.56$, $SD = 1.67$, $p = .308$) were rated as comparably useful.

4.7.3 Discussion

In Experiment 2 we partly replicated the findings of Experiment 1: In line with Hypothesis 1a, we again obtained a significant main effect of strategy instruction. In line with Hypothesis 1b, but different from Experiment 1, we found a significant main effect of feedback, likely because we provided more elaborated feedback than in Experiment 1. However, we found no instruction-by-feedback interaction (Hypothesis 2). Together, the findings suggested additive effects, as strategy instruction (rather accounted for far transfer)

as well as elaborated feedback (rather accounted for near transfer) improved students' learning outcome.

Regarding the perceived usefulness, it made no difference whether students received the combination of strategy instruction and feedback or only one of them. For the single support methods, however, we found that providing only strategy instruction without feedback led to an increased perceived usefulness of the learning phase and providing only feedback without prior strategy instruction led to an increased perceived usefulness of the practice phase. One possible explanation for why the interaction effect did not occur could be that the elaborated feedback was perceived as more useful than simple corrective feedback in Experiment 1 (Kuklick & Lindner, 2023), and thus even when students responded incorrectly (negative valence), the additional specific information provided by the elaborated feedback led from motivated feedback disengagement to more motivated feedback engagement (Grundmann et al., 2021).

4.8 Experiment 3

Given that Experiment 2 showed additive effects of combining strategy instruction and elaborated feedback, the aim of Experiment 3 was to test whether the findings could be replicated when strategy instruction and feedback were provided in a different sequence or whether the additive effects of the combination occurred only when students first go through the learning phase followed by the practice phase. Related studies which examined effects of instruction and practice phases showed that findings might depend on the sequence in which they are provided (see for related evidence: Kant et al., 2017; Kapur, 2008; Loibl et al., 2017; Sinha & Kapur, 2021; van Gog et al., 2008, 2011; van Harsel et al., 2019). In Experiment 3, we therefore contrasted the regular sequence of Experiment 1 and 2 (learning phase with strategy instruction followed by practice phase including feedback) with a flipped

Experiment 3

sequence (practice phase including feedback followed by learning phase with strategy instruction).

We stated the following hypotheses: First, we hypothesized that students in both sequences (flipped and regular) would outperform students in a control condition in which neither strategy instruction nor feedback was provided (Hypothesis 1). Second, we hypothesized that the flipped sequence (practice phase with feedback → learning phase with instruction) would result in higher learning outcomes than the regular sequence (learning phase with instruction → practice phase with feedback; Hypothesis 2).

4.8.1 Method

4.8.1.1. *Participants*

We conducted an a priori power analysis for contrast analysis using G*Power, expecting small to medium effects around $f = 0.27$ (based on Darabi et al., 2018; and related to the findings on problem solving followed by instruction by Sinha & Kapur, 2021; and research on sequencing effects in the context of worked examples by van Gog et al., 2008; see pre-registration for more details, <https://aspredicted.org/399mp.pdf>). With a set power ($1 - \beta$) of 80% ($\alpha = .05$, number of covariates = 1), we required a sample sizes of $N = 199$. Considering potential dropouts, we aimed to collect data from 240 students.

In total, 244 students from non-physics study programs of a German university participated in Experiment 3. We applied the identical exclusion criteria of Experiment 2. We excluded students who were enrolled in physics, engineering, science and technology, or technology ($n = 22$), had taken physics, technology and science, or technology as a major subject in high school ($n = 26$), all whose native language (or language level C1) was not German ($n = 0$), or who already participated in the previous experiments ($n = 9$). Furthermore, following the same procedure as in Experiment 1 and 2, to obtain robust estimates, we only included those participants who solved both attention checks correctly (n

= 18 failed at least one of the attention checks). Since the dropout was larger than expected, the final sample comprised $N = 166$ students.

With the acquired sample size, we would be able to detect medium effects of $f = .22$, which was still within the range of effect sizes reported in previous related research (see pre-registration, <https://aspredicted.org/3099mp.pdf>). Thus, we considered the acquired sample size to be acceptable.

The mean age of the students was 23.42 years ($SD = 4.62$) and 72% were female. The students were on average in their 7.10 semester ($SD = 4.48$) and most of them were enrolled in humanities ($n = 97$).

4.8.1.2. Design and Materials

In contrast to the previous experiments, we used a one-factorial design, and the participants were randomly assigned to one of the three conditions: regular sequence (learning phase with instruction → practice phase with feedback; $n = 60$), flipped sequence (practice phase with feedback → learning phase with instruction; $n = 50$), and a control condition which did not receive instruction and feedback ($n = 56$).

We used the same material as in Experiment 2.

4.8.1.3. Procedure

We followed the same procedure as in Experiment 1 and 2. As an exception, students in the flipped condition first answered the practice tasks and received feedback on their performance. Afterwards, they were provided with the strategy instruction. The regular sequence was comparable to the combined sequence of Experiment 2.

4.8.1.4. Data analysis

We applied planned contrast analyses to test our hypotheses. To investigate whether there was a general effect of the experimental conditions (H1), we contrasted the experimental

Experiment 3

conditions (flipped and regular sequence) against the control condition by using the following contrast weights: control condition: -2 ; regular sequence: 1 ; flipped sequence: 1 . To test whether the flipped sequence additionally resulted in higher learning outcome compared to the regular sequence (H2), we used the following contrast weights: control condition: 0 ; regular sequence: -1 ; flipped sequence: 1 . Furthermore, we controlled for students' prior knowledge in both contrast analyses.

4.8.2 Results

4.8.2.1. *Preliminary Analyses*

Initial boxplot analyses indicated no extreme outliers. Correlations between prior knowledge, the dependent variables and the mediation variables are presented in Appendix G. Students were comparable regarding their academic self-concept, $F(2, 163) = 1.88, p = .156, \eta_p^2 = .02$, intrinsic motivation, $F < 1$, baseline arousal, $F < 1$, baseline pleasure, $F < 1$, baseline monitoring accuracy, $F < 1$, prior knowledge, $F < 1$, and time-on-task, $F < 1$. Means and standard deviations are shown in Table 12.

Table 12*Means and Standard Deviations for all Measurements in Experiment 3*

Variables	Without instruction, without feedback (<i>n</i> = 56)	Regular sequence (learning phase – practice phase) (<i>n</i> = 60)	Flipped sequence (practice phase – learning phase) (<i>n</i> = 50)
Pretest			
Prior knowledge (0–1)	0.61 (0.23)	0.60 (0.25)	0.59 (0.24)
Practice phase			
Practice tasks (0–1)	0.45 (0.43)	0.58 (0.40)	0.61 (0.41)
Posttest			
Transfer tasks (0–1)	0.49 (0.31)	0.59 (0.30)	0.60 (0.29)
Near transfer tasks	0.59 (0.42)	0.73 (0.41)	0.70 (0.38)
Far transfer tasks	0.39 (0.34)	0.45 (0.33)	0.50 (0.32)
Mediator variables			
Mental effort (after instruction) (1–9)	5.71 (1.8)	5.67 (2.07)	5.46 (1.96)
Mental effort (after practice phase) (1–9)	6.18 (1.71)	5.37 (2.14)	5.52 (1.95)
Subjective difficulty (after instruction) (1–9)	4.61 (2.05)	5.75 (2.27)	5.30 (2.58)
Subjective difficulty (after practice phase) (1–9)	6.79 (2.09)	5.63 (2.67)	6.10 (2.38)
Monitoring accuracy (baseline) (–100%–100%)	–30.52 (29.72)	–28.19 (25.95)	–26.58 (31.30)
Monitoring accuracy practice (–100%–100%)	–8.55 (41.78)	–21.77 (35.29)	–20.92 (36.07)
Monitoring accuracy transfer (–100%–100%)	–17.41 (36.92)	–18.17 (33.25)	–27.32 (30.17)
Arousal (baseline) (1–9)	4.73 (2.03)	4.55 (1.73)	4.78 (1.82)
Arousal (after instruction) (1–9)	4.61 (1.91)	4.68 (1.79)	4.94 (1.97)
Arousal (after practice phase) (1–9)	4.64 (2.06)	4.90 (2.01)	5.32 (1.72)
Pleasure (baseline) (1–9)	5.91 (1.76)	5.82 (1.48)	5.82 (1.61)
Pleasure (after instruction) (1–9)	5.82 (1.61)	5.52 (1.56)	5.36 (1.84)
Pleasure (after practice phase) (1–9)	5.62 (1.78)	5.55 (1.86)	5.46 (1.63)

Experiment 3

Variables	Without instruction, without feedback (<i>n</i> = 56)	Regular sequence (learning phase – practice phase) (<i>n</i> = 60)	Flipped sequence (practice phase – learning phase) (<i>n</i> = 50)
Control variables			
Utility of the learning phase (1–9)	6.27 (1.42)	6.23 (1.29)	5.98 (2.00)
Utility of the practice phase (1–9)	4.62 (1.90)	5.75 (1.71)	5.18 (1.80)
Self-concept (1–4)	2.12 (0.62)	1.99 (0.71)	2.23 (0.63)
Intrinsic motivation (1–4)	2.26 (0.60)	2.13 (0.69)	2.26 (0.73)
Time in total (in seconds)	980.73 (364.89)	1033.35 (385.73)	1031.06 (357.54)

4.8.2.2. Performance on the Practice Tasks

Coherent to Experiment 1 and 2, we ran generalized mixed effect models with participants as random factor, contrasted condition as fixed factor and students' performance on the practice tasks as the dependent variable nested in participants (and repeated because of two scores on practice tasks). In addition, we controlled for students' prior knowledge.

Regarding students' performance in the practice tasks, the multilevel contrast analyses showed that students in the experimental conditions outperformed students in the control condition (see Table 12 for descriptive statistics of all measures; see Table 13 for the entire test statistics). However, there were no significant differences between the flipped sequence and the regular sequence (see Table 13)⁶.

⁶ The results of the multilevel analyses were confirmed by the a priori contrast analyses (see Appendix H).

Experiment 3

Table 13*Results of the Multilevel Contrast Analysis in Experiment 3*

Variables	Estimate	SE_b	p
Practice tasks			
Control vs. experimental condition (H1)	1.96	0.81	.016
Flipped vs. regular condition (H2)	.24	0.45	.585
Transfer tasks (overall)			
Control vs. experimental condition (H1)	.68	0.22	.002
Flipped vs. regular condition (H2)	.39	0.22	.071
Near transfer			
Control vs. experimental condition (H1)	1.38	0.68	.042
Flipped vs. regular condition (H2)	.05	0.43	.911
Far transfer			
Control vs. experimental condition (H1)	.76	0.50	.129
Flipped vs. regular condition (H2)	.23	0.29	.431

Note. Significant p -values are highlighted in bold letters. For H1, we used the following contrast weights: control condition: -2 ; regular sequence: 1 ; flipped sequence: 1 . For H2, we used the following contrast weights: control condition: 0 ; regular sequence: -1 ; flipped sequence: 1 . Furthermore, we included students' prior knowledge as covariate in both contrast analyses.

4.8.2.3. *Performance on the Transfer Tasks*

We ran again generalized mixed effect models with participants as random factor, contrasted condition as fixed factor and now students' performance on the transfer tasks as the dependent variable nested in participants (and repeated because of four scores on transfer tasks). Regarding Hypothesis 1, the multilevel contrast analyses showed that students in the experimental conditions outperformed students in the control condition regarding transfer performance (see Table 12 for descriptive statistics; see Table 13 for the entire test statistics). Regarding Hypothesis 2, there were no differences between the flipped and the regular sequence (see Table 13).

4.8.2.4. *Additional Explorative Analyses*

Similar to the previous experiments, we conducted additional explorative separate analyses for near and far transfer performance to identify any differences in near and far transfer performance. We followed the same procedure of multilevel contrast analysis but first with students' performance in the near transfer tasks as the dependent variable and in a second separate multilevel contrast analysis students' performance in the far transfer tasks as the dependent variable.

Regarding near transfer performance, the multilevel contrast analyses showed that students in the experimental conditions ($M = 0.72$, $SD = 0.39$) outperformed students in the control condition ($M = 0.59$, $SD = 0.42$; see Table 13). Regarding far transfer performance, results regarding the first contrast revealed no significant effect (see Table 13; experimental conditions: $M = 0.47$, $SD = 0.32$; control condition: $M = 0.39$, $SD = 0.34$). There were no differences between the flipped and the regular sequence for near transfer (see Table 13; flipped: $M = 0.70$, $SD = 0.38$; regular: $M = 0.73$, $SD = 0.41$), and far transfer (see Table 13; flipped: $M = 0.50$, $SD = 0.32$, regular: $M = 0.45$, $SD = 0.33$).

We conducted an additional ANOVA to explore potential differences between the regular and the flipped sequence regarding the valence of feedback (Hypothesis 2). The contrast-coded condition was the independent variable, and the valence of feedback was the dependent variable. Results showed no significant effect, $F(1, 108) = 0.21, p = .651, \eta_p^2 < .01$.

4.8.3 Discussion

In Experiment 3, we investigated whether the sequence of strategy instruction and feedback would account for reducing versus additive effects on learning. We first hypothesized that both sequences (flipped and regular) led to higher learning outcome than the control condition (Hypothesis 1). The first hypothesis was confirmed as students in the experimental conditions outperformed students in the control condition. This effect was mainly driven via near transfer performance. Second, we hypothesized that students in the flipped sequence condition outperform students in the regular sequence condition (Hypothesis 2). In contrast to our assumption, our second hypothesis was not confirmed as the flipped sequence did not contribute to higher learning outcome than the regular sequence. This pattern suggests that the obtained additive findings of elaborated feedback and strategy instruction are independent of the sequence of these instructional interventions.

4.9 General Discussion

The main aim of the present study was to investigate potential additive versus reducing effects of combining strategy instruction and feedback on learning. Furthermore, we aimed at generalizing our findings to different types of feedback (i.e., corrective vs. elaborated feedback; Experiment 1 and 2), and sequences of strategy instruction and feedback (Experiment 3). In addition, we aimed at exploratively investigating the underlying mechanisms of a potential instruction-by-feedback interaction effect.

4.9.1 Summary of Evidence

Our findings suggest that, whether combining strategy instruction and feedback results in reducing or additive effects depends on the type of feedback (corrective vs. elaborated feedback) and the type of learning goal (near vs. far transfer).

A reducing effect (i.e., an instruction-by-feedback interaction) only emerged when corrective feedback was used, and only with regard to higher-order learning outcome of far transfer (Experiment 1; see also Mertens et al., 2022). Our exploratory mediation analyses suggested that this effect was mediated by the level of arousal, but only for students who received negative feedback (moderated mediation; Kuklick & Lindner, 2023; Laudel & Narciss, 2023).

Contrarily to Experiment 1, when elaborated feedback was provided (Experiment 2), our findings showed that combining strategy instruction with elaborated feedback led to additive effects and supported students' learning. A potential explanation of how the reducing effect was counteracted could be that a) students found the elaborated feedback to be useful, as indicated by the significant effect of the perceived usefulness of the practice phase in Experiment 2 and b) that the negative emotions (i.e., arousal), when negative feedback was reported, were reduced (Grundmann et al., 2021; Kuklick & Lindner, 2023). Higher levels of usefulness may have turned motivated feedback disengagement into motivated feedback engagement (Grundmann et al., 2021; Kuklick & Lindner, 2023; Laudel & Narciss, 2023; Narciss, 1999). These findings of Experiment 2 highlight the effectiveness of elaborated feedback (Mertens et al., 2022; Shute, 2008; Wisniewski et al., 2020) not only in isolation, but also in orchestration, that is the combination with other instructional interventions, such as strategy instruction.

We assumed that another possibility to prevent motivated feedback disengagement (Grundmann et al., 2021) and to enhance feedback processing (Narciss, 2008) could be to provide the feedback at the beginning of the learning unit. Students might then be more

motivated to process the feedback, perceive it as more useful, and improve their subjective representation of standards, competencies, and task requirements (Narciss, 2013). Similar to the productive failure approach (Kapur, 2008; Sinha & Kapur, 2021), students would first work through the problems themselves and then, in a subsequent learning phase, validate and deepen their knowledge schemes, and close any knowledge gaps. Based on research on sequence effects (e.g., problem solving prior instruction; Kant et al., 2017; Loibl et al., 2017; van Gog et al., 2011), we reversed the learning and practice phases in Experiment 3. Contrary to our assumptions, there were no differences between the regular and the flipped sequence in terms of motivation, usefulness of the learning and practice phases, or learning outcomes. Experiment 3 suggested that the sequence does not matter as both sequences supported learning similarly well.

The findings of our three experiments extend previous studies on combination effects (Fyfe & Rittle-Johnson, 2016; Salden et al., 2010; Wischgoll, 2017), as they provide distinct information about when and why combining strategy instruction and formative feedback may result in additive versus reducing effects and allow to generalize the findings across several feedback implementations.

4.9.2 Theoretical Implications

What are the broader contributions of our study? The ITFL-model by Narciss (2008, 2013) proposed a reciprocal interplay between internal (cognitive, metacognitive, and affective-motivational processes) and external (e.g., feedback source, instructional context) factors that may affect the effectiveness of feedback (see also e.g., Panadero & Lipnevich, 2022; Winne & Butler, 1994, for related assumptions).

Regarding the internal factors, as a first contribution, we want to highlight the integrative character of our studies. Still to date, a large proportion of instructional research has relied on investigating cognitive mechanisms or motivational processes in isolation, whereas

affective-motivational processes have been considered only to a limited extent (see Azevedo, 2015; Braithwaite et al., 2013; D’Mello & Graesser, 2012; Goldin, 2017; Hoogerheide et al., 2019; Narciss et al., 2014, for exceptions). In our study, we integrated different perspectives (i.e., cognitive, metacognitive, affective-motivational) in three well-powered online-experimental studies comprising more than 900 students to draw conclusions about potential mechanisms of our intervention. The integration of different cognitive, metacognitive and motivational outcomes allowed us to disentangle and model the various processes which may account for potential interactions.

Moreover, we want to note that the tested explanatory mechanisms were derived from theory (Baars et al., 2013; Grundmann et al., 2021; Narciss, 2008; Sweller et al., 2011; see also pre-registration). However, we have to note that our findings rely only on self-reported measures (i.e., cognitive load, monitoring accuracy, and arousal), as we realized three online experiments. Yet it is unclear, whether and how these measures reflect the proposed theoretical constructs (see Pekrun & Bühner, 2014; Tan et al., 2021, for a critical discussion of self-reports). Replications are needed that additionally integrate less obtrusive measures, such as electrodermal activity or eye-tracking as potential behavioral proxies of (meta-)cognitive and affective-motivational measurements (e.g., Hoogerheide et al., 2019; Potter & Bolls, 2011; S. Schneider et al., 2019).

As a further contribution, regarding the external factors, our study showed that the orchestration of feedback with other instructional strategies plays an important role. We refer to the fact that we deliberately investigated combination effects of two different interventions, that is strategy instruction and computer-based feedback. In educational settings and online learning environments, strategy instruction and feedback are often combined (Magliaro et al., 2005; Rosenshine, 2008), however, most studies often relied on investigating only a single instructional intervention and research investigating to what extent both interventions affect each other’s effectiveness to enhance students’ learning is

rare and produced heterogeneous findings (Fyfe & Rittle-Johnson, 2016; Salden et al., 2010; Wischgoll, 2017). Thus, our study extends prior research as we are contributing to the scarce evidence of combining instructional interventions (Fyfe & Rittle-Johnson, 2016; Salden et al., 2010; Wischgoll, 2017; see also e.g., Murphy & Alexander, 2005).

Furthermore, our findings support prior research by highlighting the role of feedback implementation (e.g., Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Narciss, 2008; Patchan et al., 2016; Shute, 2008; Strobl et al., 2019). The design of feedback, such as the level of elaboration, may not only impact the effectiveness of feedback per se, but also might influence the effectiveness of combining instructional interventions.

Summarized, our main contribution is that our findings support the theoretical assumption (Narciss, 2008, 2017; Narciss et al., 2014; cf. Butler & Winne, 1995; Hattie & Timperley, 2007; Hayes, 2012; Shirah & Sidney, 2023) that the effectiveness of feedback depends on internal as well as on external factors: The results of our experiments indicate that whether combining strategy instruction and feedback may result in non-additive/even reducing or additive effects, strongly depends on individual student-related (internal) factors (i.e., task performance/valence, arousal) but also on feedback-related (external) factors (design decisions, i.e., elaboration level, or instructional context, i.e., orchestration with strategy instruction), probably explained by utility.

4.9.3 Practical Implications

Our findings also have implications for educational practice, as they show that combining strategy instruction and computer-based feedback is only effective under certain conditions and should therefore be used wisely. Our results showed that all types of support (strategy instruction, feedback, combination) improve learners' performance compared to learners without any instructional interventions. Contrary to the prevalent assumption, however, double support (combination of strategy instruction and feedback) does not necessarily

increase learners' performance, as demonstrated by Experiment 1. The findings counteract the commonly held belief that combining instructional interventions, which are regarded to be effective strategies in isolation, contributes to learning. Instead, practitioners should implement elaborated feedback. Our experiments have shown that elaborated feedback can have positive or additive effects on learning both in isolation (Mertens et al., 2022; Shute, 2008) and orchestrated in combination with strategy instruction (cf. Salden et al., 2010).

4.9.4 Limitations and Future Research

We would like to note that all detected effects in all three experiments were rather small. Regarding corrective feedback, this is in line with previous research which showed that corrective feedback usually produces small effects (e.g., Swart et al., 2019; van der Kleij et al., 2015; see also Mertens et al., 2022, for an overview). However, regarding elaborated feedback, previous research mostly documented medium effect sizes (see Mertens et al., 2022; Wisniewski et al., 2020, for meta-analytic evidence). It is unclear why our implemented elaborated feedback only produced small but not medium effects. One explanation refers to the fact, that we realized our study as an online experiment. Online experiments have the risk of less control and inattentive behavior by our participants, as compared to laboratory experiments (Hauser & Schwarz, 2016; Oppenheimer et al., 2009). The lower levels of control and attention may result in lower effect sizes, as compared to traditional laboratory studies (see R. S. Ryan et al., 2013).

Another limitation refers to the number of practice tasks. We implemented only two practice tasks which were realized as a multiple-choice question format. Such practice phases can be regarded to be prototypical for online learning phases. Nevertheless, implementing more challenging learning materials with more practice tasks and more transfer tasks (cf. van Harsel et al., 2020) preferably in combination with open-ended questions could help to further generalize the obtained findings.

Given that our study provides a complex picture of combining instructional interventions and the use of online experiments, we see the need to further investigate these combination effects in practice to replicate the experiments in more authentic settings. Since our findings are limited to one specific topic in physics and the instruction phase and the feedback phase were relatively short (approximately 10–15 minutes each), it is essential to replicate the findings in different study contexts, with different types of interventions, and with longer durations. Thus, further field-oriented work such as field experiments are required to replicate the findings and to increase the ecological validity of our findings (Renkl, 2013).

Our findings cannot be generalized regarding learning in general. However, we achieved our goal to provide deeper insights of the impact of combining strategy instruction with computer-based feedback on learning in an online setting.

4.9.5 Conclusions

All in all, the experiments presented in this paper provide important evidence on combining strategy instruction and feedback: The findings illustrate that combining strategy instruction and feedback does not necessarily contribute to learning, as the effectiveness depends on the provided type of feedback and the learning goals. Therefore, they provide important information for fostering cognitive skill acquisition, as they show that not only the design of instructional interventions but also their orchestration is critical to contribute to the effectiveness of digital learning environments.

5 GENERAL DISCUSSION

Learning is a broad term that is defined differently in different disciplines and dependent on the theory in which it is embedded. In this dissertation, learning in a formal context is defined with regard to the learning process that includes setting a learning goal, receiving new information, individually acquiring and processing information, connecting the new information with prior knowledge, and acting (e.g., solving a task). The change in performance after learning (i.e., the learning product) is the learning outcome. To measure this learning outcome, the level of change from before learning to after the learning process is evaluated. In addition, when investigating learning, it is important to consider the internal cognitive, metacognitive, and affective-motivational processes that are taking place during this complex process (Renkl, 2008). Cognitive processes occur during knowledge acquisition while processing into different types of memory (Atkinson & Shiffrin, 1968; Renkl, 2009). Metacognitive processes occur when learners are regulating and monitoring their actions and understanding during the learning process (Kuklick et al., 2023; Labuhn et al., 2010; Narciss, 2008, 2013). Affective-motivational processes can have an implicit or explicit impact on how a learner learn (Kuklick & Lindner, 2021, 2023; Mory, 2004; Narciss, 2008, 2013).

In order to support students' learning, feedback is considered a beneficial strategy (Kluger & DeNisi, 1996; Shute, 2008; Wisniewski et al., 2020). Educational technologies are also considered helpful supplements for instructors and teachers, as they allow for the provision of immediate feedback, including in individual learning phases, in a time-saving way (Lipnevich & Panadero, 2021; Narciss, 2013). However, previous research has provided heterogeneous findings regarding the effectiveness of computer-based feedback at fostering students' learning. Various boundary conditions have been discussed to explain the inconsistent findings, as they might have a mediating impact on the effect of computer-based feedback on learning (Burkhart et al., 2020; Fleckenstein, Liebenow, & Meyer, 2023;

GENERAL DISCUSSION

Kuklick & Lindner, 2023; Mertens et al., 2022; Narciss, 2008, 2013; Patchan et al., 2016). For example, Mertens et al. (2022) assumed that the effectiveness of computer-based feedback in the context of writing depends on the level of elaboration of that feedback. Kuklick and Lindner (2023) suggested that valence and affective processes while learners process feedback are crucial determinants of the effectiveness of computer-based feedback. Fyfe and Rittle-Johnson (2016) and Shirah and Sidney (2023) provided initial evidence that students' prior knowledge and the orchestration of feedback must be considered when implementing computer-based feedback.

In addition, previous theoretical feedback models included several potential influencing factors regarding the implementation of feedback (e.g., Narciss, 2008; Narciss & Huth, 2004; Panadero & Lipnevich, 2022). However, these models either did not mention explicit feedback-related factors (e.g., specificity, representation, elaboration, complexity) or summarized learner-related factors under a category such as students' characteristics or neglected context-related factors, such as feedbacks' orchestration with other instructional strategies. Previous models focused mainly on the processing of feedback itself instead of considering how the feedback is embedded in the learning process. However, research on the orchestration of feedback is important to take into account, as the findings showed that, for instance, providing prior instruction before providing the feedback can influence the effect of feedback (Fyfe & Rittle-Johnson, 2016; Shirah & Sidney, 2023; Wagner et al., 2024).

For practitioners, computer-based technologies seem to be a useful supplement for facilitating provision of individualized, immediate feedback to students. However, the heterogeneous findings on the effectiveness of computer-based feedback and the open questions of how computer-based feedback should be used to improve student learning, as well as which influencing factors need to be considered when choosing a feedback tool or

formulating a feedback message, make it difficult for practitioners to use computer-based feedback in practice.

Consequently, from a theoretical perspective, there was a need for a feedback model particular with regard to the implementation and orchestration of computer-based feedback in the learning process. From a practical perspective, there was a need to gain new insights into which internal and external influencing factors of the effectiveness of computer-based feedback are crucial in given subject domains, as well as which possible uses or implementation options effectively support students' learning.

Thus far, experimental studies have mostly examined the impact of different types of feedback, for example, corrective versus elaborated feedback, or only the impact of prior knowledge on feedback. Research investigating further specific internal (e.g., affective processes) and external (e.g., representation of feedback) influencing factors is still scarce (see e.g., Kuklick & Lindner, 2021, 2023, for expectations). In addition, there was an explicit need to investigate context-related boundary conditions regarding the combination of computer-based feedback with prior instruction; there has been only limited research into this topic, but initial evidence implied that the orchestration of feedback has a crucial impact on feedback's effectiveness (see e.g., Fyfe & Rittle-Johnson, 2016; McLaren et al., 2008; Salden et al., 2010; Shirah & Sidney, 2023; Wischgoll, 2017).

Thus, the overarching goal of this dissertation was to investigate the effectiveness of computer-based feedback on learning by applying a multidimensional perspective (Shute, 2008), considering feedback-related, as well as learner-related and context-related factors together in a holistic understanding. The research question was crafted as follows: What are the factors that significantly influence the effect of computer-based feedback on learning in different subject domains?

The biggest strengths of this dissertation are that computer-based feedback was examined from different perspectives and that the effects of several feedback-, learner-, and context-

related influencing factors on the effectiveness of computer-based feedback were empirically investigated. Another strength of this dissertation is that it utilized various methodological approaches (i.e., theoretical perspective, meta-analysis, empirical experiments) to pursue the goal. In this way, the robustness of the results was strengthened, as some reinforcing results were found regardless of the subject domain, tasks, procedure of providing feedback, research method, or statistical approach used to analyze the data. A third outstanding feature of the dissertation is the consideration of the combination of computer-based feedback with prior strategy instruction as there has been very little targeted research on this topic so far. Through three experiments, I have contributed substantially to the research on the interaction of these two important instructional strategies.

5.1 Summary of Evidence

In order to achieve the defined aim of this dissertation, I took a multidimensional approach by examining at the research question from theoretical and empirical perspectives in the domains of writing and physics. In the empirical studies, I applied various methods to investigate the effectiveness of computer-based feedback in fostering learning in diverse ways. In total, I have completed three articles for publication in international journals; these are included in this dissertation.

In the first paper, I developed a conceptual framework to explore the capabilities of digital media, specifically the use of computer-based feedback to foster expository writing skills. I carried out an analysis of the literature and previous research on computer-based feedback technologies to support students' expository writing to construct an overview of how computer-based feedback has been realized. Thereby, I focused on the comparison of three different types of evidence-based feedback tools and on which potential factors of computer-based feedback might have an impact on its effectiveness on learning. Through this analysis, I came to the conclusion that computer-based feedback has the potential to be an effective

and evidence-based instructional strategy to support students in their writing processes. However, I identified a need for further exploration of computer-based feedback with regard to additional characteristics of the feedback and students' prior knowledge. In addition, I identified a need for concrete research on the implementation and orchestration of computer-based feedback in order to draw direct conclusions for application in teaching contexts (see also Fleckenstein, Liebenow, & Meyer, 2023; Strobl et al., 2019).

In Paper 2, I followed the desiderata to investigate other computer-based feedback tools by conducting a systematic literature review and a meta-analysis to show what kinds of feedback tools exist to support students' expository writing and to investigate their overall effect on students' writing quality. In addition, I empirically examined the impact of feedback-related potential influencing factors which were identified as relevant in Paper 1 and, as a learner-related factor, students' prior knowledge. The results showed that computer-based feedback was not effective per se compared to control groups that did not receive feedback. Instead, the effectiveness of computer-based feedback improving students' writing quality depended on the boundary conditions. The findings of the meta-analysis revealed that learners profited more from the provision of only lower-level text quality feedback (i.e., spelling, grammar, word count, etc.) than from the provision of only on higher-level feedback (i.e., style, cohesion, organization, etc.) or from feedback on both levels. When the goal of feedback is to support students implementing aspects of higher-level text quality or on different levels of text quality and to subsequently improve their overall writing quality to a greater extent (McNamara et al., 2013; Patchan et al., 2016), students' need additional instructional support. In this regard, the analyses suggested that providing prior instruction could help students to more easily process larger amount of more complex information. Furthermore, the results of the review showed that specific feedback was more effective than general feedback and that students with lower prior knowledge profited more from the feedback. Concluding the implications of Paper 2, a question

remained: Will providing students with prior instruction on the topic of the task support them in processing more complex feedback?

In Paper 3, therefore, I aimed to investigate the effectiveness of computer-based feedback combined with strategy instruction on students' physics learning by conducting three online experiments. Besides the feedback-related influencing factors, I also investigated more learner-related factors in the experiments. In addition, I put the greatest focus on the context in which computer-based feedback is orchestrated. Therefore, I manipulated the presence of strategy instruction prior to or after the feedback. In Experiment 1, I focused on a combination of prior strategy instruction and corrective computer-based feedback. In Experiment 2, I varied the prior strategy instruction and elaborated the computer-based feedback. In Experiment 3, I investigated the sequence in which instruction and computer-based feedback were implemented in the learning process. The findings of the experiments showed that, in combination with strategy instruction, only elaborated feedback had an additive effect on students' physics learning. However, it only had an effect on near transfer tasks (lower level of complexity). Combining strategy instruction and corrective feedback had an impact on students' arousal when they answered the tasks incorrectly and could even influence students' performance negatively. Therefore, the sequence of strategy instruction and feedback was not found to be significant.

In summary, with these three papers, I have carried out research on different internal and external factors that determine the effectiveness of feedback and brought them together in a holistic understanding. I have shed light on the effects of computer-based feedback from different perspectives by applying computer-based feedback in different subject domains, examining system-generated (meta-analysis in Chapter 3) as well as person-mediated (online experiments in Chapter 4) computer-based feedback, investigating the topic theoretically as well as empirically, and using various methodological approaches (i.e., literature overview, meta-analysis, online-experiments). Based on the literature and previous research, in Paper

1, I identified potentially relevant impact factors, which were mostly related to aspects of feedback. In Paper 2, I examined the effect of these feedback-related factors as well as the impact of prior knowledge. In Paper 3, I focused on additional feedback-related factors, as well as on learner-related and context-related influencing factors. In summary, the results of this dissertation show that computer-based feedback could be a beneficial strategy for fostering learning. However, the studies also showed that successfully improving learning depends on feedback-, context-, and learner-related influencing factors. Overall, the specificity, elaboration, and level of complexity of the feedback were all found to have an effect, as did (the induction of) students' prior knowledge.

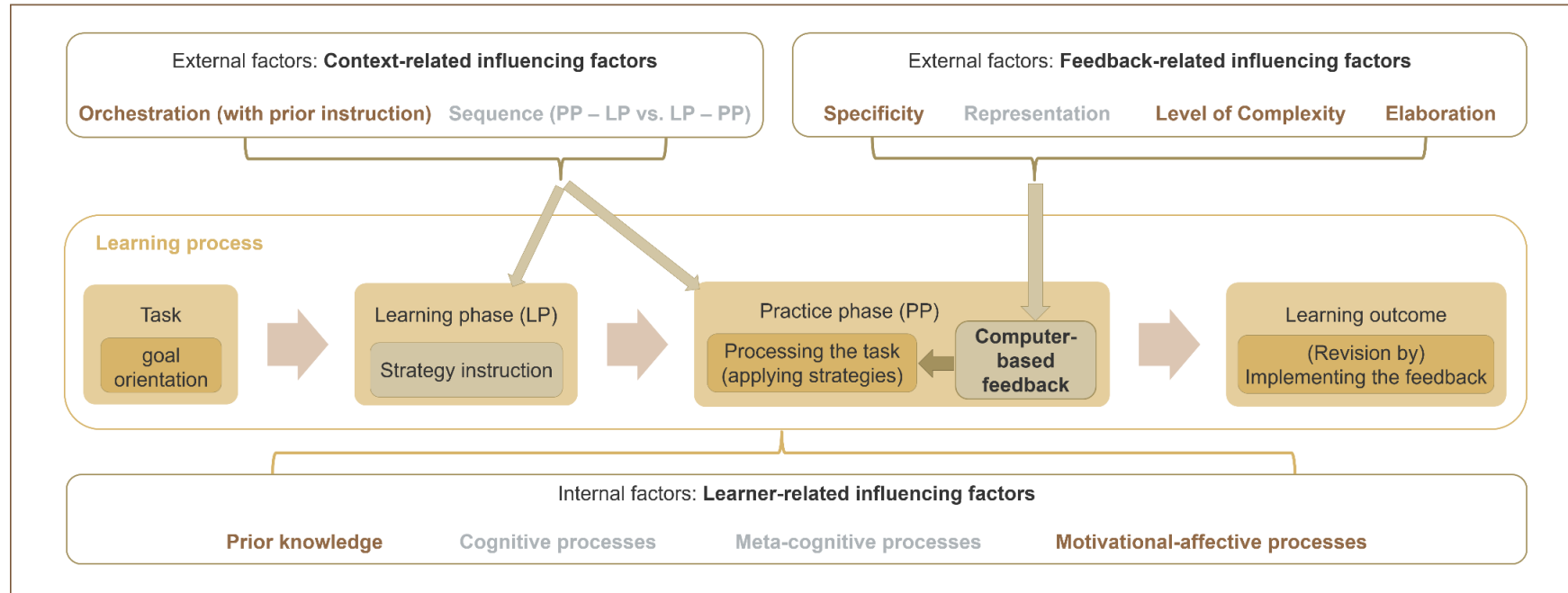
5.2 Theoretical Implications: Revised CAFCALI Model

From a theoretical perspective, different feedback models have conceptualized feedback in different ways and argued that various factors might be relevant for the effect of feedback. Since past feedback models either focused on feedback in general or did not include all these concrete computer-based feedback-, learner-, and context-related influencing factors when examining computer-based feedback embedded in students' learning processes, I have made an important contribution to the theoretical field by developing and testing the CAFCALI model using various theoretical and methodological approaches.

In this dissertation, I examined the impact of a variety of potential influencing factors. The research findings revealed that three of the four feedback-related factors, one of the two context-related factors, and two of the four learner-related factors identified in previous research had a significant impact and were capable of changing the effect of computer-based feedback. Based on these findings, I revised the CAFCALI model accordingly (see Figure 12).

Figure 12

Revised CAFCALI Model



Note. The potential influencing factors written in gray (i.e., sequence, representation, cognitive processes, metacognitive processes) were found to have no significance.

The revised CAFCALI model now shows all factors that had an impact (in brown font; the factors in gray font had no impact) on the effect of computer-based feedback: specificity of feedback, elaboration level of feedback, level of complexity, learners' prior knowledge and affective-motivational processes, and the orchestration of feedback with prior instruction (see Figure 12). Accordingly, whether computer-based feedback fosters learning depends on feedback-, context-, and learner-related influencing factors. This test of the CAFCALI model, has contributed new evidence about how computer-based feedback can foster learning and highlighted what practitioners and researchers should keep in mind when implementing computer-based feedback.

5.2.1 Indications Regarding Orchestration

Through this dissertation, the limited research that exists on the orchestration of computer-based feedback was replicated and extended. I was able to replicate the non-additive effect of strategy instruction and corrective feedback found by Fyfe and Rittle-Johnson (2016) and confirm the results of Shirah and Sidney (2023), indicating that inducing relevant prior knowledge prior to additional feedback does not lead to additional benefits for learners. Furthermore, I was able to extend previous research by showing that providing elaborated feedback can counteract this interaction effect and achieve additive effects. Since Fyfe and Rittle-Johnson (2016) conducted their research in the domain of mathematics with primary school children, and Shirah and Sidney (2023) focused on the domain of writing with undergraduate students, I was able to contribute new insights regarding the orchestration of computer-based feedback by implementing my experiments in the subject context of physics. Overall, this strengthens the assumption that this initial evidence regarding the feedback-instruction-interaction can be found across subject domains and educational levels.

5.2.2 Indications Regarding Specificity

Specificity of feedback describes the level of detail with which a feedback message is formulated (Goodman et al., 2004; Patchan et al., 2016): generic (e.g., in the form of a numerical score or a grade at the end of a school year) or specific (e.g., analytic feedback in the form of a text-based message that includes feedback comments regarding different aspects of the learner's performance). The findings of this dissertation confirm previous assumptions, that the more specific a feedback message, the more effective it is (see also Hattie & Timperley, 2007; Lachner et al., 2017a; Patchan et al., 2016; Shute, 2008; Wisniewski et al., 2020). When interpreting these findings, however, it must be noted that the effect of elaboration may outshine this effect of specificity, as elaborated feedback can be viewed as more specific and analytical than corrective feedback, which is often very generic (Mertens et al., 2022; Shute, 2008).

5.2.3 Indications Regarding the Level of Complexity

Regarding the level of complexity, this dissertation found that feedback on a hierarchically higher level (e.g., cohesion, organization; see Chapters 2 and 3) and that feedback with the aim of supporting higher-order student learning outcomes (i.e., far transfer; see Chapter 4) was more difficult for the learners to implement. This is also in line with previous research, which found that feedback in the context of more complex tasks or on multiple cues (e.g., global cohesion, error analysis) revealed either no positive effects on learning or weaker effects than those of simpler feedback on a lower level of complexity (Kluger & DeNisi, 1996; Kulhavy et al., 1985; Shute, 2008). Processing and implementing more complex feedback information or feedback information about a more complex learning objective is more demanding (Sweller et al., 1998) and requires more prerequisites (de Jong & Ferguson-Hessler, 1996; Renkl, 2009). For example, in order to write a globally cohesive text, the learner first needs to know how an argumentative text should be structured. This

requires rhetorical prior knowledge as well as a higher level of self-regulation, as no detailed steps on how to proceed can be specified; instead, the feedback information must be interpreted and then implemented in an adapted way (Renkl, 2008). For higher-order learning outcomes, it is not enough for learners to reproduce information they have read before, as in recall tasks. Instead, transfer requires understanding the content, which therefore requires deeper processing and also a higher level of self-regulation in order to implement knowledge transfer (Baartman & de Bruijn, 2011; de Jong & Ferguson-Hessler, 1996).

5.2.4 Indications Regarding Elaboration

In line with previous research, elaborated computer-based feedback proved to be more effective than corrective feedback (e.g., Bangert-Drowns et al., 1991; Kluger & DeNisi, 1996; Mertens et al., 2022; Narciss, 2008; Narciss & Huth, 2004; Shute, 2008). In addition, new evidence was provided, as elaborated computer-based feedback in combination with prior knowledge instruction also proved to be more effective than corrective feedback. Corrective feedback provides the information whether students solved the task correctly or not (knowledge of result; Narciss, 2012) and in the case of knowledge of correct response feedback, additionally, which would have been the correct answer (Narciss, 2012). In contrast, elaborated feedback messages address specific aspects of students' responses that are related to the task and provide additional information on what was wrong and, more importantly, on how to solve the task correctly (Kluger & DeNisi, 1996; Mertens et al., 2022). The goal-oriented nature of elaborated feedback, which aims to help learners develop learning strategies in a self-regulated way so that they may solve future tasks correctly, can explain why elaborated feedback has a more positive effect than less goal-oriented and action-oriented corrective feedback.

5.2.5 Indications Regrading Prior Knowledge

Previous research has consistently shown that prior knowledge has a decisive influence on learning and thus on the processing of feedback information (Bryson et al., 1991; Flower et al., 1992; Fyfe & Rittle-Johnson, 2016; Kellogg, 2008; Krause et al., 2009; Mory, 2004; Narciss & Huth, 2004; Patchan et al., 2016). There is also a large amount of evidence suggesting that feedback is particularly suitable for learners with less prior knowledge (Allen et al., 2016; Fyfe & Rittle-Johnson, 2016; Krause et al., 2009; Strobl et al., 2019). This pattern also appeared in the research results of this dissertation. Overall, I found that feedback is more effective for learners with a lower level of prior knowledge than for those with a higher level of prior knowledge. These results support prior research that states that low-achieving students, in particular, need additional support in order to reach learning goals (Castro-Alonso et al., 2021). In contrast, high-achieving students are more likely to apply self-regulated and adequate strategies (Castro-Alonso et al., 2021), and they herefore do not benefit as much from corrective feedback in particular, as this provides them with little new information (Mertens et al., 2022). In contrast, elaborated feedback proved to be effective for learners with either level of prior knowledge, as it provides more information on how the respective task can be solved (see Mertens et al., 2022, for related findings).

5.2.6 Indications Regarding Arousal

Regarding learners' affective-motivational processes, this dissertation found that learners' arousal (dependent on valence) seems to have an effect. In line with previous research, I found that when students receive negative corrective feedback, it has a negative effect on their performance due to their arousal (see also Kluger & DeNisi, 1996; Kuklick & Lindner, 2023, for related evidence). A possible explanation for this detrimental effect could be that the learners perceive the feedback on their self level (cf. Hattie & Timperley, 2007). To increase feedback perception and focus it on the process level, Laudel and Narciss

(2023) recommended providing a short introduction in advance to enhance feedback acceptance. In line with Kuklick and Lindner (2023), the results of this dissertation suggest that providing elaborated feedback, including more concrete information on what was wrong and why and how the learner can improve the performance, can counteract the negative mediating effect of arousal on students' performance after receiving negative feedback. This is also in line with previous findings, which showed that too general or unclear feedback can hinder learning and frustrate learners (Kluger & DeNisi, 1996; Moreno, 2004; Shute, 2008; Williams, 1997). By providing more elaborated feedback that includes more helpful information for the learner, it seems that this frustration effect could be counteracted.

5.2.7 Neglecting Factors: Sequence, Representation, and Cognitive and Metacognitive Processes

Besides the several crucial factors, there were also four factors (spanning the feedback-related, learner-related, and context-related categories) that revealed no significance in the research of this dissertation.

The sequence of the orchestration of feedback refers to the time point at which computer-based feedback is provided: at the beginning of the learning process (e.g., practice phase followed by learning phase) or, in the case of formative feedback, in the middle of the learning process (e.g., starting the learning process with a learning phase that is then followed by a practice phase that includes feedback). To my knowledge, to date, no research has specifically examined the sequence of strategy instruction and computer-based feedback. Research on the sequence of the learning phase (i.e., instruction or worked examples) and the practice phase (i.e., problem solving) has provided contradictory findings, arguing either in favor of providing a practice phase followed by a learning phase (Kapur, 2012; Loibl et al., 2017; Sinha & Kapur, 2021) or in favor of providing instruction followed by a practice phase (Kant et al., 2017; van Gog et al., 2011). So, there are no references against which to compare in order to determine if my finding that the sequence does not matter is an artifact

or if when the feedback is provided really makes no difference to the effectiveness of computer-based feedback. Bayesian methods are necessary to investigate this further. It must also be noted that I implemented elaborated feedback when manipulating the sequence (see Chapter 4.8, Experiment 3). Consequently, it may be that, when implementing a different type of feedback (e.g., corrective feedback), the sequence would reveal significance. Elaborated feedback counteracted the non-additive instruction-by-feedback interaction effect in Experiment 2 (see Chapter 4.7) in the regular sequence and might have also outshone the sequence effect here.

To date, there has been little research investigating the visual representation of feedback messages. Interestingly, there is a connection to the factor of specificity, as empirical evidence has shown that holistic scores are less effective in fostering learning than specific comments (Butler, 1987; Kluger & DeNisi, 1996). Since the findings of this dissertation confirmed this effect by showing that specificity had a significant impact on the effectiveness of computer-based feedback, it is surprising that representation did not become significant. However, it must also be emphasized that differences in the representation formats were visible in the moderator analysis of studies with pretest-posttest designs in the meta-analysis (see Chapter 3.8.5). However, this effect disappeared when a control group was included. In addition, it must be noted that only 24 effect sizes with pretest-posttest comparisons and 15 effect sizes with pretest-posttest-control group designs were included in the meta-analysis. Among these comparisons, there was relatively little variation in the visualization of feedback messages because the same feedback tools were used in many studies. Hence, it is possible that no moderating effect of the representation factor occurred due to the small sample size.

One explanation for why cognitive internal processes did not appear to have a significant impact on the effectiveness of computer-based feedback in Experiment 1 (see Chapter 4.6) could be that the feedback message was relatively short and rarely repeated the content of

the prior instruction. In Experiment 2, the content from the instruction and the feedback were indeed doubled, but in the case of the feedback message, the content was placed in a practical context and therefore provided additional information for the learners. For this reason, the addition of feedback information to the prior instruction may not have been overwhelming but may have rather increased the germane load (Paas & van Gog, 2006). In addition, when presenting the feedback, we made sure that the task, the learner's answer, and the feedback were all presented close to each other to avoid a split attention effect (Chandler & Sweller, 1991; Sweller et al., 2011). It is possible that this counteracted cognitive overload (Sweller et al., 2011). Overall, it can be interpreted as a positive result that it does not seem to matter which feedback type is implemented or how computer-based feedback is orchestrated; the internal cognitive processes did not influence the effectiveness of the feedback. However, since mental effort and perceived difficulty were assessed only on the basis of self-assessments in the research of this dissertation, it would be interesting to obtain more information about the way in which feedback is processed using physiological measurement instruments such as eye tracking.

Previous research regarding the effect of feedback on metacognitive processes (students' judgments of learning) has indicated that feedback can lead to a more accurate self-assessment by students (e.g., Kuklick et al., 2023; Labuhn et al., 2010; Nietfeld et al., 2006). By visualizing students' current performance through the feedback and providing them with information that enables them to compare their performance with the target performance, one would assume that the feedback would lead to an improvement in self-assessment (Kuklick et al., 2023; Labuhn et al., 2010). In combination with prior instruction, the assumption was that students felt confident and indeed overestimated their performance, which influenced how they regulated subsequent learning processes (Baars et al., 2013; Finn & Metcalfe, 2014). Surprisingly, in the research of this dissertation, metacognitive accuracy was not significantly improved by feedback and thus did not have a decisive influence on

the learners' performance (see Chapter 4). The descriptive results showed tendencies that support the assumptions of previous research that feedback led to a more accurate self-assessment by students. Contrary to expectations (see Baars et al., 2013; Dunning, 2011; Finn & Metcalfe, 2014; Kruger & Dunning, 1999, for related evidence on students overestimating their competence due to overconfidence), however, the learners assessed themselves quite accurately at the beginning of the learning process, which is why the feedback could not lead to any major improvement and the effect was therefore not significant.

5.3 Future Research to Further Refine the CAFCALI Model

To sum up, the results of this dissertation show that, in line with previous research, the elaboration and specificity of feedback, the level of complexity, and learners' prior knowledge are the main influencing factors regarding the effectiveness of computer-based feedback; these proved to be crucial, regardless of subject domain, type of feedback generation (system-generated or person-mediated), or the approach used to test significance. However, there are still some open questions.

That learners with lower prior knowledge profited more from computer-based feedback than did students with higher prior knowledge begs the following question: How can high-performing learners be adequately supported? One recommendation by Shute (2008) is to provide high-achieving students with delayed feedback. Other possibilities include the implementation of adaptive ITS or innovative technologies such as AI-supported systems, which could offer valuable potential (e.g., large language models, such as GPT or BERT; Montenegro-Rueda et al., 2023; Scheurer et al., 2022; Tsai et al., 2020).

Regarding the second learner-related factor that revealed significance, it must be noted that arousal was measured by self-assessments due to the online study setting of this dissertation. Self-reports are discussed critically (see, e.g., Pekrun & Böhner, 2014; Tan et

al., 2021), and previous research has indicated that learners struggle to self-assess their performance or competences adequately and tend to overestimate their performance (Dunning, 2011; Kruger & Dunning, 1999). Therefore, further research in which arousal is measured using more objective measurement methods would help to clarify these results. The aim of my future research is therefore to replicate and expand these results by collecting and analyzing physiological data on learners' electrodermal activity (EDA; measured by an EDA wristband; see Hoogerheide et al., 2019, for an example). The significant effects of underlying mechanisms and feedback traits raise further questions about other potential influencing factors that would be worthwhile to investigate, including acceptance of digital media, impact of social presence, feedback receptivity, or ratings of computer versus human feedback. It would also be interesting to explore the influence of different domains, as I found similar or overlapping results regarding the effectiveness and influencing factors of computer-based feedback in the domain of expository writing and in the context of physical problem solving. However, these subjects were not specifically manipulated in this dissertation and were examined separately.

Regarding the feedback-related factors, future research is needed to investigate how more complex learning could be supported using computer-based feedback and how computer-based feedback messages should be formulated so that they contain all the necessary information but avoid unnecessary complexity. Mason and Bruning (2001) reported that the level of complexity produced contradictory findings on learning depending on the interaction with other factors (e.g., students' prior knowledge, time when the feedback is provided, level of elaboration, specificity of feedback). For example, Mason and Bruning (2001) reported that delayed feedback is more recommended for learners with higher achievement, especially for more complex tasks, whereas students with lower achievement profited more from immediate feedback regardless of the level of complexity. In addition, one can argue that elaborated feedback is more complex than corrective feedback because it addresses

multiple cues, whereas corrective feedback simply tells the learner if the answer was correct or not. However, research suggests that elaborated feedback is more effective than corrective feedback because of the extra information on how to improve their performance and correct potential misconceptions (Mertens et al., 2022; Shute, 2008). In order to increase the selectivity of individual factors, further research is needed. For example, Patchan et al. (2016) assumed that the localization of feedback information (where it is displayed in the learning environment) is another feedback-related factor connected to the representation and complexity of the feedback message (see also Goodman & Wood, 2004; Goodman et al., 2011, for related evidence). For example, in the text, as with highlighting, or on another page, as can be the case with concept map feedback or text-based comments. Research based on the cognitive load theory indicates that it makes a difference whether related information is integrated or presented separately (Chandler & Sweller, 1991; Sweller et al., 1998, 2011). In this dissertation, unfortunately, it is unclear whether the localization of the feedback influenced the effect of representation or specificity, as localization was not mentioned or explained transparently in most of the primary studies in the meta-analysis (Chapter 3). In order to examine the effect of representation as well as the other feedback-related factors in more detail, implementing more open scientific practices would be helpful (e.g., a more transparent description of the individual constructs being investigated would allow for more precise analyses).

Regarding the context-related factors, it would be interesting to investigate how computer-based feedback (in combination with prior instruction) in the field and especially in a school context affects the writing and problem solving performance of students, since the experiments in this dissertation were all realized in an online setting with university students. However, comparing the results of the experiments (see Chapter 4) with previous research provides initial evidence that the pattern identified in this dissertation in the domain of physics can be replicated in primary school (Fyfe & Rittle-Johnson, 2016), with different

intervention materials and in different subject domains, as, for instance, Fyfe and Rittle-Johnson (2016) produced similar results in the domain of mathematics and Shirah and Sidney (2023) found related evidence in the domain of writing.

5.4 Practical Implications

When learners face difficulties during the learning process and need additional support, feedback is considered an effective strategy to improve learning. However, providing individualized and immediate feedback is very time-consuming and difficult, particularly in heterogeneous courses. To relieve instructors, the question investigated in this dissertation was under which circumstances computer-based feedback could be used to improve students' learning. Practitioners may ask themselves what type of computer-based feedback they should use or what they need to be aware of, in terms of possible influencing factors, when using computer-based feedback.

5.4.1 Conscious Choice of a Feedback Tool

In general, it can be said that computer-based feedback must be handled with caution and that one cannot simply use any tool for every aim of learning (see also Mertens et al., 2022; van der Kleij et al., 2015). My meta-analysis (see Chapter 3) has shown that there is great heterogeneity in computer-based feedback tools and that not every tool is equally effective. When selecting the tools, or when deciding to use person-mediated computer-based feedback (where the teacher can create the feedback messages and choose the visualization, but the computer is the medium through which the feedback is immediately and simultaneously communicated), the first implication for practice is the importance of implementing computer-based feedback in an elaborated and specific way. General or non-elaborated feedback was consistently shown to be less or not at all effective in fostering learning. The combination of non-elaborated feedback with prior strategy instruction can even have a reducing effect compared to a single intervention. In contrast, elaborated

feedback was consistently shown to be effective (see also Bangert-Drowns et al., 1991; Mertens et al., 2022), both as a single intervention and in combination with prior strategy instruction. Accordingly, practitioners should formulate feedback messages as simply and precisely as possible, providing not only information about the correctness of the answer or the error but also hints, suggestions, or prompts on how to improve performance related to the goal of the task, repetition of the strategy that is needed to solve the task, or provision of the steps needed to solve the task correctly (Kluger & DeNisi, 1996; Mertens et al., 2022; Narciss & Huth, 2004; Shute, 2008). However, educators must be careful not to create a message that is too complex.

5.4.2 Conscious Choice of Complexity and Orchestration

This dissertation also showed that the effectiveness of both system-generated feedback from writing tools and person-mediated computer-based feedback (in the context of physics) depend on the level of complexity. In line with previous research, the findings of the dissertation consistently showed that computer-based feedback is more effective at a lower level of complexity and that more support is needed for more complex feedback (i.e., higher level of text quality, far transfer tasks). Practitioners must therefore be careful when using feedback at a hierarchically higher level or for the purpose of improving a learning outcome at a hierarchically higher level (see Mertens et al., 2022, for similar tendencies). In this case, it is recommended that the instructors monitor more closely whether learners are benefiting from the feedback or are perhaps overwhelmed or affectively frustrated and need further support.

Regarding the orchestration of computer-based feedback, it can be assumed that there is usually some kind of instruction (in the learning phase) at the beginning of every learning process, followed by a practice phase with feedback (Kang et al., 2023). However, in previous research, the effect of feedback was mostly investigated without considering the

feedback environment and potential synergy- or interaction effects with prior instruction (Kang et al., 2023). The findings of this dissertation, however, show the importance of considering whether instruction is provided prior to computer-based feedback. Replicating initial evidence on the orchestration of instruction and computer-based feedback, the findings of this dissertation indicate that the induction of (topic-relevant) prior knowledge through prior strategy instruction in combination with corrective feedback has no additional added value for learners (see also Fyfe & Rittle-Johnson, 2016; Shirah & Sidney, 2023, for similar results). However, this interaction effect can be counteracted with elaborated feedback (see Chapter 4, Experiment 2), as combining prior instruction and additional elaborated feedback produced additive effects.

For practice, the results imply first, that more is not always better (Kang et al., 2023): If you have to work with corrective feedback, provide only one type of support (i.e., instruction or corrective feedback). But in daily practice, there are two problems with this suggestion: On the one hand, it is often not possible to omit the instruction, and on the other hand, corrective computer-based feedback alone has no or only a small effect (Kuklick & Lindner, 2023; Mertens et al., 2022; Narciss, 2013; Shute, 2008). Therefore, it is recommended that practitioners use the second option and apply a combination of prior instruction and elaborated computer-based feedback, as this orchestration had positive effects.

In addition, when we take the prior knowledge of the learners into account, the findings of this dissertation showed that, in line with previous research, computer-based corrective feedback tends to be more beneficial for learners with a lower level of prior knowledge (Fyfe & Rittle-Johnson, 2016; Mertens et al., 2022; cf. expertise-reversal effect, Kalyuga, 2007). Elaborated feedback was shown to be beneficial for both learners with lower and learners with higher levels of prior knowledge (see also Mertens et al., 2022, for similar results). This is another reason why it is recommended for practitioners to implement elaborated feedback, as no group of learners will be detrimentally affected. Although elaborated feedback is more

time-consuming to provide than corrective feedback, this dissertation's results show that the additive effect of prior instruction and elaborated feedback (independent of learners' levels of prior knowledge) also occurs when implementing person-mediated computer-based feedback. This means that teachers can formulate the feedback message in advance and the computer system can provide it to the students immediately and simultaneously.

5.5 Conclusion

This dissertation systematically investigated the effectiveness of computer-based feedback on learning and its boundary conditions from different perspectives. The special feature here was that several feedback-, learner-, and context-related factors influencing the effect of computer-based feedback were empirically examined, which has never been done before in this way.

The newly developed CAFCALI model summarizes the theoretically conceptualized and empirically tested influencing factors of computer-based feedback and thus provides groundbreaking information for researchers and practitioners. First, future research could refer to this model and expand it to include further possible influencing factors. However, when investigating computer-based feedback, it is important that scientists are aware that the effectiveness of computer-based feedback is higher when the feedback is specific and elaborated and that learners with less prior knowledge benefit more from feedback, especially on a hierarchically lower level or when a learning outcome is targeted at a lower order. In addition, the instructional context must be taken into account, as providing instruction prior to the feedback determines and distorts the effect of the feedback type (a reducing effect when corrective feedback is provided and an additive effect when elaborated feedback is provided).

Second, practitioners can also use the CAFCALI model as a guide and use the evidence provided when creating feedback messages or selecting feedback tools. In essence,

computer-based feedback has a learning-enhancing effect when it is elaborated and specific. This dissertation has shown various ways in which digital media and educational technologies can be used to implement elaborated feedback in a time-saving manner.

Finally, the goal of identifying how computer-based feedback can be effective was achieved by showing that the effectiveness of such feedback depends on feedback-, learner-, and context-related influencing factors. If these are taken into account, computer-based feedback can be a helpful and time-saving supplement in teaching and in teachers' instruction and support. For practitioners, it can also be utilized to provide individual, immediate feedback simultaneously, and they can use the saved time to provide increased support for individual learners.

In summary, computer-based feedback can be a useful instructional strategy for supporting learners during their learning processes in a digital environment that is worth exploring and implementing further as long as the influencing factors are carefully studied and considered.

References

Articles included in the meta-analyses (Chapter 3) are marked with an asterisk.

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>

Ainley, M., Corrigan, M., & Richardson, N. (2005). Students, tasks and emotions:

Identifying the contribution of emotions to students' reading of popular culture and popular science texts. *Learning and Instruction, 15*(5), 433–477.

<https://doi.org/10.1016/j.learninstruc.2005.07.011>

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with

multiple representations. *Learning and Instruction, 16*(3), 183–198.

<https://doi.org/10.1016/j.learninstruc.2006.03.001>

Alharbi, W. (2017). E-feedback as a scaffolding teaching strategy in the online language

classroom. *Journal of Educational Technology Systems, 46*(2), 239–251.

<https://doi.org/10.1177/0047239517697966>

Allen, L. K., Jacovina, M. E., & McNamara, D. S. (2016). Computer-based writing

instruction. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.). *Handbook of writing research* (2nd ed., pp. 316–329). Guilford.

Alruwais, N., Wills, G., & Wald, M. (2018). Advantages and challenges of using e-assessment.

International Journal of Information and Education Technology, 8(1), 34–37.

<https://doi.org/10.18178/ijiet.2018.8.1.1008>

Anderson, J. R., & Schunn, C. D. (2000). Implications of the ACT-R learning theory: No

magic bullets. In R. Glaser (Ed.), *Advances in instructional psychology: Educational design and cognitive science* (Vol. 5, pp. 1–34). Lawrence Erlbaum Associates.

- Arnsten, A. F. T. (2009). Stress signalling pathways that impair prefrontal cortex structure and function. *Nature Reviews Neuroscience*, *10*(6), 410–422.
<https://doi.org/10.1038/nrn2648>
- Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research*, *70*(2), 181–214. <https://doi.org/10.3102/00346543070002181>
- Atkinson, R. K., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *Psychology of Learning and Motivation*, *2*, 89–195.
[https://doi.org/10.1016/S0079-7421\(08\)60422-3](https://doi.org/10.1016/S0079-7421(08)60422-3)
- Attali, Y. (2004). Exploring the feedback and revision features of Criterion. *Journal of Second Language Writing*, *14*(3), 191–205.
- Attali, Y., & Burstein, J. (2006). Automated essay scoring with e-rater® V.2. *Journal of Technology, Learning, and Assessment*, *4*(3).
- Azevedo, R. (2015). Defining and measuring engagement and learning in science: Conceptual, theoretical, methodological, and analytical issues. *Educational Psychologist*, *50*(1), 84–94. <https://doi.org/10.1080/00461520.2015.1004069>
- Azevedo, R., & Bernard, R. M. (1995). A meta-analysis of the effects of feedback in computer-based instruction. *Journal of Educational Computing Research*, *13*(2), 111–127.
<https://doi.org/10.2190/9LMD-3U28-3A0G-FTQ>
- Baars, M., Visser, S., van Gog, T., de Bruin, A., & Paas, F. (2013). Completion of partially worked-out examples as a generation strategy for improving monitoring accuracy. *Contemporary Educational Psychology*, *38*(4), 395–406.
<https://doi.org/10.1016/j.cedpsych.2013.09.001>
- Baartman, L. K., & de Bruijn, E. (2011). Integrating knowledge, skills and attitudes: Conceptualising learning processes towards vocational competence. *Educational Research Review*, *6*, 125–134. <https://doi.org/10.1016/J.EDUREV.2011.03.001>

References

- Baddeley, A. (1992). Working memory. *Science*, 255(5044), 556–559.
<https://doi.org/10.1126/science.1736359>
- Bai, S., Hew, K. F., & Huang, B. (2020). Does gamification improve student learning outcome? Evidence from a meta-analysis and synthesis of qualitative data in educational contexts. *Educational Research Review*, 30, 100322.
<http://doi.org/10.1016/j.edurev.2020.100322>
- Balcazar, F. E., Hopkins, B. L., & Suarez, Y. (1985). A critical, objective review of performance feedback. *Journal of Organizational Behavior Management*, 7(3–4), 65–89. https://doi.org/10.1300/J075V07N03_05
- Bangert-Drowns, R. L., Kulik, C.-L. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, 61(2), 213–238. <https://doi.org/10.3102/00346543061002213>
- Barbieri, C. A., Miller-Cotto, D., Clerjoste, S. N., & Chawla, K. (2023). A meta-analysis of the worked examples effect on mathematics performance. *Educational Psychology Review*, 35(1), 11. <https://doi.org/10.1007/s10648-023-09745-1>
- Becker, B. J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical and Statistical Psychology*, 41(2), 257–278.
<https://doi.org/10.1111/j.2044-8317.1988.tb00901.x>
- Becker-Mrotzek, M., Grabowski, J., Jost, J., Knopp, M., & Linnemann, M. (2014). Adressatenorientierung und Kohärenzherstellung im Text [Audience focus and generating coherence in the text]. *Didaktik Deutsch*, 19(37), 21–43.
<https://doi.org/10.25656/01:17151>
- Becker-Mrotzek, M., & Schindler, K. (2007). Texte schreiben [Writing texts]. In H. Günther, U. Bredel, & M. Becker-Mrotzek (Hrsg.), *Kölner Beiträge zur Sprachdidaktik* [Cologne contributions to language didactics] (series A, pp. 7–26). Gilles & Francke Verlag.

- Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Lawrence Erlbaum Associates, Inc.
- Berninger, V. W., & Swanson, H. L. (1994). Modification of the Hayes and Flower model to explain beginning and developing writing. In E. Butterfield (Ed.), *Advances in Cognition and Educational Practice* (Vol. 2, pp. 57–82). JAI Press.
- 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>
- Billett, S. (2001). Knowing in practice: Re-conceptualising vocational expertise. *Learning and Instruction*, *11*(2001), 431–452.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7–74.
<https://doi.org/10.1080/0969595980050102>
- Block, J. (1995). A contrarian view of the five-factor approach to personality description. *Psychological Bulletin*, *117*(2), 187–215. <https://doi.org/10.1037/0033-2909.117.2.187>
- Boksem, M. A. S., Meijman, T. F., & Lorist, M. M. (2005). Effects of mental fatigue on attention: An ERP study. *Cognitive Brain Research*, *25*(1), 107–16.
<https://doi.org/10.1016/j.cogbrainres.2005.04.011>
- Boksem, M. A. S., Meijman, T. F., & Lorist, M. M. (2006). Mental fatigue, motivation and action monitoring. *Biological Psychology*, *72*(2), 123–32.
<https://doi.org/10.1016/j.biopsycho.2005.08.007>
- Bono, R., Alarcón, R., & Blanca, M. J. (2021). Report quality of generalized linear mixed models in psychology: A systematic review. *Frontiers in Psychology*, *12*(666182).
<https://doi.org/10.3389/fpsyg.2021.666182>
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons.

References

- Braithwaite, J. J., Watson, D. G., Jones, R., & Rowe, M. (2013). A guide for analysing electrodermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments. *Psychophysiology*, *49*(1), 1017–1034.
- Brindle, M., Graham, S., Harris, K. R., & Herbert, M. (2016). Third and fourth grade teacher's classroom practices in writing: A national survey. *Reading and Writing*, *29*, 929–954. <https://doi.org/10.1007/s11145-015-9604-x>
- Bromme, R., Jucks, R., & Rambow, R. (2004). Experten-Laien-Kommunikation im Wissensmanagement [Communication of experts and novices in knowledge management]. In G. Reinmann & H. Mandl (Eds.). *Der Mensch im Wissensmanagement: Psychologische Konzepte zum besseren Verständnis und Umgang mit Wissen* [The human being in knowledge management: Psychological concepts for a better understanding and handling of knowledge] (pp. 176–188). Hogrefe.
- Brooks, C., Burton, R., van der Kleij, F., Ablaza, C., Carroll, A., Hattie, J., & Neill, S. (2021). Teachers activating learners: The effects of a student-centred feedback approach on writing achievement. *Teaching and Teacher Education*, *105*, 103387. <https://doi.org/10.1016/j.tate.2021.103387>
- Brooks, C., Carroll, A., Gillies, R. M., & Hattie, J. (2019). A matrix of feedback for learning. *Australian Journal of Teacher Education*, *44*(4), 14–32. <http://dx.doi.org/10.14221/ajte.2018v44n4.2>
- Bryson, M., Bereiter, C., Scardamalia, M., & Joram, E. (1991). Going beyond the problem as given: Problem solving in expert and novice writers. In R. J. Sternberg & P. A. Frensch (Eds.), *Complex problem solving: Principles and mechanisms* (pp. 61–84). Lawrence Erlbaum Associates, Inc.
- Burgers, C., Eden, A., Van Engelenburg, M. D., & Buningh, S. (2015). How feedback boosts motivation and play in a brain-training game. *Computers in Human Behavior*, *48*, 94–103. <https://doi.org/10.1016/j.chb.2015.01.038>

- *Burkhart, C., Lachner, A., & Nückles, M. (2020). Using spatial contiguity and signaling to optimize visual feedback on students' written explanations. *Journal of Educational Psychology*. <https://doi.org/10.1037/edu0000607>
- Butler, D., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, 65(3), 245–281. <https://doi.org/10.2307/1170684>
- Butler, R. (1987). Task-involving and ego-involving properties of evaluation: Effects of different feedback conditions on motivational perceptions, interest, and performance. *Journal of Educational Psychology*, 79(4), 474–482. <https://doi.org/10.1037/0022-0663.79.4.474>
- Carroll, J. M., & Kay, D. S. (1988). Prompting, feedback and error correction in the design of a scenario machine. *International Journal of Man-Machine Studies*, 28, 11–27.
- 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, 1379–1407. <https://doi.org/10.1007/s10648-021-09606-9>
- Chan, A. W., & Altman, D. G. (2005). Epidemiology and reporting of randomised trials published in PubMed Journals. *The Lancet*, 365, 1159–1162. [http://dx.doi.org/10.1016/S0140-6736\(05\)71879-1](http://dx.doi.org/10.1016/S0140-6736(05)71879-1)
- Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition & Instruction*, 8(4), 293–332.
- Chanquoy, L. (2001). How to make it easier for children to revise their writing? A study of text revision from 3rd to 5th grades. *British Journal of Educational Psychology*, 71, 15–41. <https://doi.org/10.1348/000709901158370>
- Chanquoy, L. (2009). Revision processes. In R. Beard, D. Myhill, J. Riley, & M. Nystrand (Eds.), *The SAGE handbook of writing development* (pp. 80–97). SAGE.

References

- Chenoweth, N., & Hayes, J. R. (2001). Fluency in writing. Generating text in L1 and L2. *Written Communication, 18*, 80–98. <https://doi.org/10.1177/0741088301018001004>
- Cho, K., & MacArthur, C. (2010). Student revision with peer and expert reviewing. *Learning and Instruction, 20*(4), 328–338. <https://doi.org/10.1016/j.learninstruc.2009.08.006>
- Cho, K., & MacArthur, C. (2011). Learning by reviewing. *Journal of Educational Psychology, 103*(1), 73–84. <https://doi.org/10.1037/a0021950>
- Chodorow, M., Gamon, M., & Tetreault, J. (2010). The utility of article and preposition error correction systems for English language learners: Feedback and assessment. *Language Testing, 27*(3), 419–436. <https://doi.org/10.1177/0265532210364391>
- Clark, R. E. (1983). Reconsidering research on Learning from media. *Review of Educational Research, 53*(4), 445–459. <https://doi.org/10.3102/00346543053004445>
- Cohen, J. (2013). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge. <https://doi.org/10.4324/9780203771587>
- Concha, S., & Paratore, J. R. (2010). Local coherence in persuasive writing: An exploration of Chilean students' metalinguistic knowledge, writing process, and writing products. *Written Communication, 28*(1), 34–69. <https://doi.org/10.1177/0741088310383383>
- Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.). (2019). *The Handbook of Research Synthesis and Meta-Analysis*. Russell Sage Foundation. <https://doi.org/10.7758/9781610448864>
- Crissman, K. (2006). *The design and utilization of effective worked examples: A meta-analysis* (Publication No. 3208114) [Doctoral dissertation, The University of Nebraska-Lincoln]. ProQuest Dissertations Publishing.
- Crossley, S. A., & McNamara, D. S. (2011). Understanding expert ratings of essay quality: Coh-Metrix analyses of first and second language writing. *International Journal of*

- Continuing Engineering Education and Life Long Learning*, 21(2-3), 170–191.
<https://doi.org/10.1504/IJCEELL.2011.040197>
- Crossley, S. A., & McNamara, D. S. (2016). Say more and be more coherent: How elaboration and cohesion can increase writing quality. *The Journal of Writing Research*, 7(3), 351–370. <https://doi.org/10.17239/jowr-2016.07.03.02>
- Crossley, S. A., Varner, L. K., Roscoe, R. D., & McNamara, D. S. (2013). Using automated indices of cohesion to evaluate an intelligent tutoring system and an automated writing evaluation system. In H. C. Lane, K. Yacef, J. Mostow, & P. Pavlik (Eds.), *Artificial intelligence in education. AIED 2013. Lecture notes in Computer Science* (Vol. 7926). Springer. https://doi.org/10.1007/978-3-642-39112-5_28
- Dai, W., Lin, J., Jin, H., Li, T., Tsai, Y.-S., Gašević, D., & Chen, G. (2023). Can large language models provide feedback to students? A case study on ChatGPT. *IEEE International Conference on Advanced Learning Technologies (ICALT)*, Orem, UT, USA, 323–325, <https://doi.org/10.1109/ICALT58122.2023.00100>
- Darabi, A., Arrington, T. L., & Sayilir, E. (2018). Learning from failure: A meta-analysis of the empirical studies. *Educational Technology Research and Development*, 66(5), 1101–1118. <https://doi.org/10.1007/s11423-018-9579-9>
- de Jong, T., & Ferguson-Hessler, M. G. (1996). Types and qualities of knowledge. *Educational Psychologist*, 31(2), 105–113.
- DeLeeuw, K. E., & Mayer, R. E. (2008). A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load. *Journal of Educational Psychology*, 100(1), 223–234. <https://doi.org/10.1037/0022-0663.100.1.223>
- Dikli, S. (2006). An overview of automated scoring of essays. *The Journal of Technology, Learning and Assessment*, 5(1). <https://ejournals.bc.edu/index.php/jtla/article/view/1640>

References

- D'Mello, S., & Graesser, A. C. (2012). Dynamics of affective states during complex learning. *Learning and Instruction, 22*(2), 145–157.
<https://doi.org/10.1016/j.learninstruc.2011.10.001>
- Duijnhouwer, H., Prins, F. J., & Stokking, K. M. (2012). Feedback providing improvement strategies and reflection on feedback use: Effects on students' writing motivation, process, and performance. *Learning and Instruction, 22*(3), 171–184.
<https://doi.org/10.1016/j.learninstruc.2011.10.003>
- Dunning, D. (2011). The Dunning–Kruger effect: On being ignorant of one's own ignorance. In M. Zanna, & J. Olsen (Eds.), *Advances in experimental social psychology* (Vol. 44, pp. 247–296). Academic Press. <https://doi.org/10.1016/B978-0-12-385522-0.00005-6>
- Duval, S. J., & Tweedie, R. L. (2000a). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association, 95*(449), 89–98. <https://doi.org/10.1080/01621459.2000.10473905>
- Duval, S. J., & Tweedie, R. L. (2000b). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*(2), 455–463. <https://doi.org/10.1111/j.0006-341x.2000.00455.x>
- Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual review of psychology, 53*(1), 109–132.
<https://doi.org/10.1146/annurev.psych.53.100901.135153>
- Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology, 61*, 101859.
<https://doi.org/10.1016/j.cedpsych.2020.101859>
- Efklikes, A., & Dina, F. (2004). Feedback from one's self and from the others: Their effect on affect. *Hellenic Journal of Psychology, 1*(2), 179–202.

- Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, *315*(7109), 629–634.
<https://doi.org/10.1136/bmj.315.7109.629>
- Eitel, A. (2016). How repeated studying and testing affects multimedia learning: Evidence for adaptation to task demands. *Learning and Instruction*, *41*, 70–84.
<https://doi.org/10.1016/j.learninstruc.2015.10.003>
- Epstein, M. L., Lazarus, A. D., Calvano, T. B., Matthews, K. A., Hendel, R. A., Epstein, B. B., & Brosvic, G. M. (2002). Immediate feedback assessment technique promotes learning and corrects inaccurate first responses. *The Psychological Record*, *52*, 187–201. <https://doi.org/10.1007/BF03395423>
- Fineout-Overholt, E., & Johnston, L. (2005). Teaching EBP: asking searchable, answerable clinical questions. *Worldviews on Evidence-Based Nursing*, *2*(3), 157–160.
<https://doi.org/10.1111/j.1741-6787.2005.00032.x>
- Finn, B. (2015). Measuring motivation in low-stakes assessments. *ETS Research Report Series*, *2*, 1–17. <https://doi.org/10.1002/ets2.12067>
- Finn, B., & Metcalfe, J. (2014). Overconfidence in children's multi-trial judgments of learning. *Learning and Instruction*, *32*, 1–9.
<https://doi.org/10.1016/j.learninstruc.2014.01.001>
- Fleckenstein, J., Liebenow, L. W., & Meyer, J. (2023). Automated feedback and writing: A multi-level meta-analysis of effects on students' performance. *Frontiers in Artificial Intelligence*, *6*, 1162454. <https://doi.org/10.3389/frai.2023.1162454>
- Fleckenstein, J., Reble, R., Meyer, J., Jansen, T., Liebenow, L. W., Möller, J., & Köller, O. (2023). Digitale Schreibförderung im Bildungskontext: Ein systematisches Review [Digital writing support in educational contexts: A systematic review]. In K. Scheiter, & I. Gogolin (Eds.), *Bildung für eine digitale Zukunft* [Education for a digital future]

References

- (Edition ZfE, Vol. 15, pp. 3–25). Springer VS. https://doi.org/10.1007/978-3-658-37895-0_1
- Flower, L., & Hayes, J. R. (1980). The dynamics of composing: Making plans and juggling constraints. In L. Gregg, & E. Steinberg (Eds.), *Cognitive processes in writing* (pp. 31–50). Erlbaum.
- Flower, L., Schriver, K. A., Carey, L., Haas, C., & Hayes, J. R. (1992). Planning in writing: The cognition of a constructive process. In S. P. Witte, N. Nakadate, & R. D. Cherry (Eds.), *A rhetoric of doing: Essays on written discourse in honor of James L. Kinneavy* (pp. 181–243). Southern Illinois University Press.
- Fong, C. J., Patall, E. A., Vasquez, A. C., & Stautberg, S. (2019). A meta-analysis of negative feedback on intrinsic motivation. *Educational Psychology Review, 31*, 121–162. <https://doi.org/10.1007/s10648-018-9446-6>
- Fong, C. J., & Schallert, D. L. (2023). “Feedback to the future”: Advancing motivational and emotional perspectives in feedback research. *Educational Psychologist, 58*(3), 146–161. <https://doi.org/10.1080/00461520.2022.2134135>
- *Frost, K. L. (2008). The effects of automated essay scoring as a high school classroom intervention [Doctoral dissertation, University of Nevada, Las Vegas]. *UNLV Retrospective Theses & Dissertations, 2839*. <http://dx.doi.org/10.25669/37p4-kdh7>
- Fyfe, E. R., & Rittle-Johnson, B. (2016). Feedback both helps and hinders learning: The causal role of prior knowledge. *Journal of Educational Psychology, 108*, 82–97. <https://doi.org/10.1037/edu0000053>
- Fyfe, E. R., & Rittle-Johnson, B. (2017). Mathematics practice without feedback: A desirable difficulty in a classroom setting. *Instructional Science, 45*(2), 177–194. <https://doi.org/10.1007/s11251-016-9401-1>
- Galesic, M. (2006). Dropouts on the web: Effects of interest and burden experienced during an online survey. *Journal of Official Statistics, 22*(2), 313–328.

- Gaspard, H., Dicke, A.-L., Flunger, B., Schreier, B., Häfner, I., Trautwein, U., & Nagengast, B. (2015). More value through greater differentiation: Gender differences in value beliefs about math. *Journal of Educational Psychology, 107*(3), 663–677. <https://doi.org/10.1037/edu0000003>
- Gaspard, H., Wigfield, A., Jiang, Y., Nagengast, B., Trautwein, U., & Marsh, H. W. (2018). Dimensional comparisons: How academic track students' achievements are related to their expectancy and value beliefs across multiple domains. *Contemporary Educational Psychology, 52*, 1–14. <https://doi.org/10.1016/j.cedpsych.2017.10.003>
- Geiser, S., & Studley, R. (2001). *UC and SAT: Predictive validity and differential impact of the SAT I and SAT II at the University of California*. University of California. https://doi.org/10.4324/9780203463932_uc_and_the_sat
- Göhlich, M., Wulf, C., & Zirfas, J. (2007). Pädagogische Zugänge zum Lernen. Eine Einleitung. [Pedagogical approaches to learning. An introduction]. In M. Göhlich, C. Wulf, & J. Zirfas (Eds.), *Pädagogische Theorien des Lernens* [Pedagogical theories of learning] (pp. 7–19). Beltz.
- Goldin, G. A. (2017). Motivating desires for classroom engagement in the learning of mathematics. In C. Andrà, D. Brunetto, E. Levenson, & P. Liljedahl (Eds.), *Teaching and learning in maths classrooms* (pp. 219–229). Springer. https://doi.org/10.1007/978-3-319-49232-2_21
- Goldman, S. R. (1989). Strategy instruction in mathematics. *Learning Disability Quarterly, 12*(1), 43–55. <https://doi.org/10.2307/1510251>
- Goleman, D. (1995). *Emotional intelligence*. Bantam Books, Inc.
- Golke, S., Dörfler, T., & Artelt, C. (2015). The impact of elaborated feedbacks on text comprehension within a computer-based assessment. *Learning and Instruction, 39*, 123–136. <https://doi.org/10.1016/j.learninstruc.2015.05.009>

References

- Goodman, J. S., & Wood, R. E. (2004). Feedback specificity, learning opportunities, and learning. *Journal of Applied Psychology, 89*(5), 809–821. <https://doi.org/10.1037/0021-9010.89.5.809>
- Goodman, J. S., Wood, R. E., & Chen, Z. (2011). Feedback specificity, information processing, and transfer of training. *Organizational Behavior and Human Decision Processes, 115*, 253-267. <https://doi.org/10.1016/j.obhdp.2011.01.001>
- Goodman, J. S., Wood, R. E., & Hendrickx, M. (2004). Feedback specificity, exploration, and learning. *Journal of Applied Psychology, 89*, 248–262. <https://doi.org/10.1037/0021-9010.89.2.248>
- Graham, S. (2019). Changing how writing is taught. *Review of Research in Education, 43*(1), 277–303. <https://doi.org/10.3102/0091732X18821125>
- Graham, S., Harris, K. R., & Hebert, M. (2011). *Informing writing: The benefits of formative assessment*. Alliance for Excellence in Education.
- Graham, S., & Herbert, M. (2011). Writing to read: A meta-analysis of the impact of writing and writing instruction on reading. *Harvard Educational Review, 81*(4), 710–744. <http://dx.doi.org/10.17763/haer.81.4.t2k0m13756113566>
- Graham, S., Hebert, M., & Harris, K. R. (2015). Formative assessment and writing: A meta-analysis. *The Elementary School Journal, 115*(4), 523–547. <https://doi.org/10.1086/681947>
- Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology, 99*(3), 445. <https://doi.org/10.1037/0022-0663.99.3.445>
- Gräsel, C., Bruhn, J., Mandl, H., & Fischer, F. (1997). Lernen mit Computernetzen aus konstruktivistischer Perspektive [Learning with computer networks from a constructivistic perspective]. *Unterrichtswissenschaft [Teaching science], 25*(1), 4–18.

- Greller, M. M., & Herold, D. M. (1975). Sources of feedback: Preliminary investigation. *Organizational Behavior and Human Performance*, *13*, 244–256.
[https://doi.org/10.1016/0030-5073\(75\)90048-3](https://doi.org/10.1016/0030-5073(75)90048-3)
- Gruber, H., Mandl, H., & Renkl, A. (2000). Was lernen wir in Schule und Hochschule: Träges Wissen? [What do we learn in school and university: Inertial knowledge?]. In H. Mandl, & J. Gerstenmaier (Eds.), *Die Kluft zwischen Wissen und Handeln* [The gap between knowledge and action] (pp. 139–156). Hogrefe.
- Grundmann, F., Scheibe, S., & Epstude, K. (2021). When ignoring negative feedback is functional: Presenting a model of motivated feedback disengagement. *Current Directions in Psychological Science*, *30*(1), 3–10.
<https://doi.org/10.1177/0963721420969386>
- Guo, R., Palmer-Brown, D., Lee, S. W., & Cai, F. F. (2014). Intelligent diagnostic feedback for online multiple-choice questions. *Artificial Intelligence Review*, *42*, 369–383. <https://doi.org/10.1007/s10462-013-9419-6>
- Hacker, D. J. (1994). Comprehension monitoring as a writing process. In E. C. Butterfield (Ed.), *Children's writing: Toward a process theory of the development of skilled writing*. JAI Press.
- Harris, K. R., Santangelo, T., & Graham, S. (2008). Self-regulated strategy development in writing: Going beyond NLEs to a more balanced approach. *Instructional Science*, *36*, 395–408. <https://doi.org/10.1007/s11251-008-9062-9>
- Harrison, C. J., Könings, K. D., Schuwirth, L., Wass, V., & Van der Vleuten, C. (2015). Barriers to the uptake and use of feedback in the context of summative assessment. *Advances in Health Sciences Education*, *20*(1), 229–245. <https://doi.org/10.1007/s10459-014-9524-6>
- Hattie, J., & Clarke, S. (2019). *Visible learning: Feedback*. Routledge.

References

- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Hauser, D. J., & Schwarz, N. (2016). Attentive turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods, 48*(1), 400–407. <https://doi.org/10.3758/s13428-015-0578-z>
- Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis. A regression-based approach*. Guilford Press.
- Hayes, J. R. (2012). Modeling and Remodeling Writing. *Written Communication, 29*(3), 369–388. <https://doi.org/10.1177/0741088312451260>
- Hayes, J. R., & Flower, L. S. (1986). Writing research and the writer. *American Psychologist, 41*(10), 1106–1113. <https://doi.org/10.1037/0003-066x.41.10.1106>
- Hayes, J. R., Flower, L., Schriver, K., Stratman, J. F., & Carey, L. (1987). Cognitive processes in revision. In S. Rosenberg (Ed.), *Reading, Writing, and Language Processes* (pp. 176–240). Cambridge University Press.
- Hayes, J. R., & Olinghouse, N. G. (2015). Can cognitive writing models inform the design of the common core state standards? *The Elementary School Journal, 115*(4), 480–497. <https://doi.org/10.1086/681909>
- Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ (Clinical research ed.), 327*(7414), 557–560. <https://doi.org/10.1136/bmj.327.7414.557>
- Hoerger, M. (2010). Participant dropout as a function of survey length in internet-mediated university studies: Implications for study design and voluntary participation in psychological research. *Cyberpsychology, Behavior, and Social Networking, 13*(6), 697–700. <https://doi.org/10.1089/cyber.2009.0445>

- Holman, L. D. (2011). *Automated writing evaluation program's effects on student writing achievement* [Doctoral dissertation, Tennessee State University, Nashville]. ProQuest LLC.
- 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*(11), 45–56.
<https://dx.doi.org/10.1037/edu0000272>
- Hoogerheide, V., & Roelle, J. (2020). Example-based learning: New theoretical perspectives and use-inspired advances to a contemporary instructional approach. *Applied Cognitive Psychology, 34*(4), 787–792. <https://doi.org/10.1002/acp.3706>
- Huitt, W., Monetti, D., & Hummel, J. (2009). Designing direct instruction. In C. Reigeluth, & A. Carr-Chellman, *Instructional-design theories and models* (Vol. 3, pp. 73–97). Mahwah.
- Ivanjek, L., Morris, L., Schubatzky, T., Hopf, M., Burde, J.-P., Haagen-Schützenhöfer, C., Dopatka, L., Spatz, V., & Wilhelm, T. (2021). Development of a two-tier instrument on simple electric circuits. *Physical Review Physics Education Research, 17*(2), 020123.
<https://doi.org/10.1103/PhysRevPhysEducRes.17.020123>
- Jackson, D., & Turner, R. (2017). Power analysis for random-effects meta-analysis. *Research Synthesis Methods, 8*(3), 290–302. <https://doi.org/10.1002/jrsm.1240>
- Jacob, L. (2022). *Learning by explaining: How implementation-and student-related boundary conditions determine the effectiveness of generating an explanation to a fictitious peer* [Doctoral dissertation, Universität Tübingen]. Tübingen Library Publishing. https://tobias-lib.uni-tuebingen.de/xmlui/bitstream/handle/10900/117896/Dissertation_Leonie%20Jacob%20Mai%202021.pdf?sequence=2&isAllowed=y

References

- 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>
- Jiang, Y., Rosenzweig, E. Q., & Gaspard, H. (2018). An expectancy-value-cost approach in predicting adolescent students' academic motivation and achievement. *Contemporary Educational Psychology, 54*, 139–152. <https://doi.org/10.1016/j.cedpsych.2018.06.005>
- Kalyuga, S. (2006). Rapid cognitive assessment of learners' knowledge structures. *Learning and Instruction, 16*, 1–11. <https://doi.org/10.1016/j.learninstruc.2005.12.002>
- Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review, 19*, 509–539.
<https://doi.org/10.1007/s10648-007-9054-3>
- Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect, *Educational Psychologist, 38*(1), 23–31. https://doi.org/10.1207/S15326985EP3801_4
- Kalyuga, S., & Renkl, A. (2010). Expertise reversal effect and its instructional implications: Introduction to the special issue. *Instructional Science, 38*, 209–215.
<https://doi.org/10.1007/s11251-009-9102-0>
- Kalyuga, S., & Sweller, J. (2004). Measuring knowledge to optimize cognitive load factors during instruction. *Journal of Educational Psychology, 96*, 558–568.
<https://doi.org/10.1037/0022-0663.96.3.558>
- Kalyuga, S., & Sweller, J. (2014). The redundancy principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 247–262). Cambridge University Press. <https://doi.org/10.1017/CBO9781139547369.013>
- Kang, Y., Ha, H., & Lee, H. S. (2023). When more is not better: Effects of interim testing and feature highlighting in natural category learning. *Educational Psychology Review, 35*, 51. <https://doi.org/10.1007/s10648-023-09772-y>

- 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>
- Kapur, M. (2008). Productive failure. *Cognition and Instruction, 26*(3), 379–424. <https://doi.org/10.1080/07370000802212669>
- Kapur, M. (2012). Productive failure in learning the concept of variance. *Instructional Science, 40*(4), 651–672. <https://doi.org/10.1007/s11251-012-9209-6>
- Kellogg, R. T. (2008). Training writing skills: A cognitive developmental perspective. *Journal of Writing Research, 1*(1), 1–26. <https://doi.org/10.17239/jowr-2008.01.01.1>
- Kellogg, R. T., & Raulerson, B. A. (2007). Improving the writing skills of college students. *Psychonomic Bulletin & Review, 14*(2), 237–242. <https://doi.org/10.3758/bf03194058>
- Kellogg, R. T., & Whiteford, A. P. (2009). Training advanced writing skills: The case for deliberate practice. *Educational Psychologist, 44*(4), 250–266. <https://doi.org/10.1080/00461520903213600>
- *Kellogg, R. T., Whiteford, A. P., & Quinlan, T. (2010). Does automated feedback help students learn to write?, *Journal of Educational Computing Research, 42*(2), 173–196. <https://doi.org/10.2190/EC.42.2.c>
- Kieft, M., Rijlaarsdam, G., & van den Bergh, H. (2008). An aptitude–treatment interaction approach to writing-to-learn. *Learning and Instruction, 18*(4), 379–390. <https://doi.org/10.1016/j.learninstruc.2007.07.004>
- Kim, M. K. (2012). Cross-validation study of methods and technologies to assess mental models in a complex problem solving situation. *Computers in Human Behavior, 28*(2), 703–717. <https://doi.org/10.1016/j.chb.2011.11.018>
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95*(2), 163–182. <https://doi.org/10.1037/0033-295X.95.2.163>

References

- 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
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254–284. <https://doi.org/10.1037/0033-2909.119.2.254>
- Kluger, A. N., Lewinsohn, S., & Aiello, J. (1994). The influence of feedback on mood: Linear effects on pleasantness and curvilinear effects on arousal. *Organizational Behavior and Human Decision Processes, 60*, 276–299.
- Kovanović, V., Gašević, D., Dawson, S., Joksimović, S., Baker, R. S., & Hatala, M. (2015). Does time-on-task matter? Implications for the validity of learning analytics findings. *Journal of Learning Analytics, 2*(3), 81–110. <http://dx.doi.org/10.18608/jla.2015.23.6>
- Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory Into Practice, 41*(4), 212–218. https://doi.org/10.1207/s15430421tip4104_2
- Krause, S., Kelly, J., Corkins, J., & Tasooji, A. (2009). The role of prior knowledge on the origin and repair of misconceptions in an introductory class on materials science and engineering. *Proceedings of the Research in Engineering Education Symposium 2009*, Palm Cove, QLD, Australia.
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121–1134. <https://doi.org/10.1037/0022-3514.77.6.1121>
- Kuechler, W., & Simkin, M. (2003). How well do multiple choice tests evaluate student understanding in computer programming classes? *Journal of Information Systems Education, 14*(4), 389–400.

- Kuklick, L., Greiff, S., & Lindner, M. A. (2023). Computer-based performance feedback: Effects of error message complexity on cognitive, metacognitive, and motivational outcomes. *Computers & Education, 200*, 104785. <https://doi.org/10.1016/j.compedu.2023.104785>
- Kuklick, L., & Lindner, M. A. (2021). Computer-based knowledge of results feedback in different delivery modes: Effects on performance, motivation, and achievement emotions. *Contemporary Educational Psychology, 67*, 102001. <https://doi.org/10.1016/j.cedpsych.2021.102001>
- Kuklick, L., & Lindner, M. A. (2023). Affective-motivational effects of performance feedback in computer-based assessment: Does error message complexity matter? *Contemporary Educational Psychology, 73*, 102146. <https://doi.org/10.1016/j.cedpsych.2022.102146>
- Kulhavy, R. W., & Stock, W. A. (1989). Feedback in written instruction: The place of response certitude. *Educational Psychology Review, 1*(4), 279–308. <https://doi.org/10.1007/bf01320096>
- Kulhavy, R. W., White, M. T., Topp, B. W., Chan, A. L., & Adams, J. (1985). Feedback complexity and corrective efficiency. *Contemporary Educational Psychology, 10*(3), 285–291. [https://doi.org/10.1016/0361-476X\(85\)90025-6](https://doi.org/10.1016/0361-476X(85)90025-6)
- Labuhn, A. S., Zimmerman, B. J., & Hasselhorn, M. (2010). Enhancing students' self-regulation and mathematics performance: The influence of feedback and self-evaluative standards. *Metacognition and Learning, 5*(2), 173–194. <https://doi.org/10.1007/s11409-010-9056-2>
- 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>

References

- *Lachner, A., Burkhart, C., & Nückles, M. (2017a). Formative computer-based feedback in the university classroom: Specific concept maps scaffold students' writing. *Computers in Human Behavior*, 72(4), 459–469. <http://dx.doi.org/10.1016/j.chb.2017.03.008>
- *Lachner, A., Burkhart, C., & Nückles, M. (2017b). Mind the gap! Automated concept map feedback supports students in writing cohesive explanations. *Journal of Experimental Psychology: Applied*, 23(1), 29–46. <http://dx.doi.org/10.1037/xap0000111>
- *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>
- Lachner, A., & Scheiter, K. (2020). Digitale Medien zur Realisierung adaptiven Unterrichts [Digital media to realize adaptive teaching]. In S. Aufenanger, B. Eickelmann, A. Feindt, & A.-M. Kamin (Eds.), *Digitale Bildung* [Digital education]. Friedrich Verlag.
- Lachner, A., Weinhuber, M., & Nückles, M. (2019). To teach or not to teach the conceptual structure of mathematics? Teachers undervalue the potential of principle-oriented explanations. *Contemporary Educational Psychology*, 58, 175–185. <https://doi.org/10.1016/j.cedpsych.2019.03.008>
- Lam, C. F., DeRue, D. S., Karam, E. P., & Hollenbeck, J. R. (2011). The impact of feedback frequency on learning and task performance: Challenging the “more is better” assumption. *Organizational Behavior and Human Decision Processes*, 116(2), 217–228. <https://doi.org/10.1016/j.obhdp.2011.05.002>
- Langner, R., Steinborn, M. B., Chatterjee, A., Sturm, W., & Willmes, K. (2010). Mental fatigue and temporal preparation in simple reaction-time performance. *Acta Psychologica*, 133(1), 64–72. <https://doi.org/10.1016/j.actpsy.2009.10.001>

- Laudel, H., & Narciss, S. (2023). The effects of internal feedback and self-compassion on the perception of negative feedback and post-feedback learning behavior. *Studies in Educational Evaluation, 77*, 101237. <https://doi.org/10.1016/j.stueduc.2023.101237>
- Lepper, M., & Gurtner, J. L. (1989). Children and computers: Approaching the twenty-first century. *American Psychologist, 44*, 170–178.
- Lipnevich, A. A., Berg, D. A. G., & Smith, J. K. (2016). Toward a model of student response to feedback. In G. T. L. Brown, & L. R. Harris (Eds.), *Handbook of human and social conditions in assessment* (pp. 169–185). Routledge.
- Lipnevich, A. A., & Panadero, E. (2021). A review of feedback models and theories: Descriptions, definitions, and conclusions. *Frontiers in Educations, 6*, 720195. <https://doi.org/10.3389/educ.2021.720195>
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Sage Publications, Inc.
- Loibl, K., Roll, I., & Rummel, N. (2017). Towards a theory of when and how problem solving followed by instruction supports learning. *Educational Psychology Review, 29*, 693–715. <https://doi.org/10.1007/s10648-016-9379-x>
- Loibl, K., & Rummel, N. (2014). The impact of guidance during problem solving prior to instruction on students' inventions and learning outcomes. *Instructional Science, 42*, 305–326. <https://doi.org/10.1007/s11251-013-9282-5>
- Lorist, M. M., Klein, M., Nieuwenhuism, S., De Jong, R., Mulder, G., & Meijman, T. F. (2000). Mental fatigue and task control: Planning and preparation. *Psychophysiology, 37*(5), 614–625. <https://doi.org/10.1017/S004857720099005X>
- MacArthur, C. A. (2015). Instruction in evaluation and revision. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (2nd ed., pp. 272–287). Guilford Press.

References

- MacArthur, C. A., & Graham, S. (2016). Writing research from a cognitive perspective. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 24–40). The Guilford Press.
- Magliaro, S. G., Lockee, B. B., & Burton, J. K. (2005). Direct instruction revisited: A key model for instructional technology. *Educational technology research and development*, 53(4), 41–55. <https://doi.org/10.1007/BF02504684>
- Mason, B. J., & Bruning, R. (2001). Providing feedback in computer-based instruction: What the research tells us. *CLASS Research Report (No. 9)*. Center for Instructional Innovation, University of Nebraska-Lincoln.
- Mautone, P. D., & Mayer, R. E. (2001). Signaling as a cognitive guide in multimedia learning. *Journal of Educational Psychology*, 93(2), 377–389. <https://doi.org/10.1037/0022-0663.93.2.377>
- Mayer, R. E. (1996). Learning strategies for making sense out of expository text: The SOI model for guiding three cognitive processes in knowledge construction. *Educational Psychology Review*, 8(4), 357–371. <https://doi.org/10.1007/BF01463939>
- *McCarthy, K. S., Roscoe, R. D., Likens, A. D., & McNamara, D. S. (2019). Checking it twice: Does adding spelling and grammar checkers improve essay quality in an automated writing tutor? In S. Isotani, E. Millán, A. Ogan, P. Hastings, B. McLaren, & R. Luckin (Eds.), *Artificial Intelligence in Education, 20th International Conference, AIED 2019 Chicago, IL, USA, June 25–29, 2019 Proceedings, Part I* (pp. 270–282). Springer. <https://doi.org/10.1007/978-3-030-23204-7>
- McDermott, L. C., & Shaffer, P. S. (1992). Research as a guide for curriculum development: An example from introductory electricity. Part I: Investigation of student understanding. *American Journal of Physics*, 60(11), 994–1003. <https://doi.org/10.1119/1.17003>

- McKenzie, J. E., Brennan, S. E., Ryan, R. E., Thomson, H. J., & Johnston, R. V. (2022). Analysing data and undertaking meta-analyses. In J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (Eds.). *Cochrane Handbook for Systematic Reviews of Interventions* (Version 6.3). Cochrane.
- McLaren, B. M., Lim, S., & Koedinger, K. R. (2008). When and how often should worked examples be given to students? New results and a summary of the current state of research. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 2176–2181). Austin, TX: Cognitive Science Society.
- McNamara, D. S., Crossley, S. A., & Roscoe, R. (2013). Natural language processing in an intelligent writing strategy tutoring system. *Behavior Research Methods*, *45*, 499–515. <https://doi.org/10.3758/s13428-012-0258-1>
- McNamara, D. S., Crossley, S. A., Roscoe, R. D., Allen, L. K., & Dai, J. (2015). A hierarchical classification approach to automated essay scoring. *Assessing Writing*, *23*, 35–59. <https://doi.org/10.1016/j.asw.2014.09.002>
- Mertens, U., Finn, B., & Lindner, M. A. (2022). Effects of computer-based feedback on lower- and higher-order learning outcomes: A network meta-analysis. *Journal of Educational Psychology*, *114*(8), 1743–1772. <https://doi.org/10.1037/edu0000764>
- Meurers, D., De Kuthy, K., Nuxoll, F., Rudzewitz, B., & Ziai, R. (2019). Scaling up intervention studies to investigate real-life foreign language learning in school. *Annual Review of Applied Linguistics*, *39*. <https://doi.org/10.1017/S0267190519000126>
- Mitchell, M. L., & Jolley, J. M. (2010). *Research design explained* (7th ed.). Wadsworth.
- Möckel, T., Beste, C., & Wascher, E. (2015). The effects of time on task in response selection – An ERP study of mental fatigue. *Scientific Reports*, *5*(1), 10113. <https://doi.org/10.1038/srep10113>

References

- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA Statement. *PLoS Med*, *6*(7). <https://doi.org/10.1371/journal.pmed.1000097>
- Möller, J. (2008). Lernmotivation [Learning motivation]. In A. Renkl (Ed.), *Lehrbuch Pädagogische Psychologie* [Manual of pedagogical psychology] (pp. 263–298). Huber.
- Montenegro-Rueda, M., Fernández-Cerero, J., Fernández-Batanero, J. M., & López-Meneses, E. (2023). Impact of the implementation of ChatGPT in education: A systematic review. *Computers*, *12*(8), 153. <https://doi.org/10.3390/computers12080153>
- Moore, N. S., & MacArthur, C. A. (2016). Student use of automated essay evaluation technology during revision. *Journal of Writing Research*, *8*.
<http://dx.doi.org/10.17239/jowr-2016.08.01.05>
- Moreno, R. (2004). Decreasing cognitive load for novice students: Effects of explanatory versus corrective feedback in discovery-based multimedia. *Instructional Science*, *32*, 99–113. <https://doi.org/10.1023/B:TRUC.0000021811.66966.1d>
- Moreno, R., & Mayer, R. E. (2002). Verbal redundancy in multimedia learning: When reading helps listening. *Journal of Educational Psychology*, *94*, 156–163.
<https://doi.org/10.1037/0022-0663.94.1.156>
- Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods*, *11*(2), 364–386.
<https://doi.org/10.1177/1094428106291059>
- Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods*, *7*(1), 105–125. <https://doi.org/10.1037//1082-989X.7.1.105>
- Mory, E. H. (2004). Feedback research revisited. In D. H. Jonassen (Ed.), *Handbook of research on educational communications and technology* (pp. 745–783). Lawrence Erlbaum.

- Murphy, P. K., & Alexander, P. A. (2005). *Understanding how students learn: A guide for instructional leaders*. Corwin Press.
- Narciss, S. (1999, April 19–23). *Motivational effects of the informativeness of feedback* [Paper presentation]. The Annual Meeting of the American Educational Research Association, Montreal, Quebec, Canada.
- Narciss, S. (2006). *Informatives tutorielles Feedback. Entwicklungs- und Evaluationsprinzipien auf der Basis instruktionspsychologischer Erkenntnisse* [Informative tutorial feedback. Development and evaluation principles based on the findings of instructional psychology]. Waxmann.
- Narciss, S. (2008). Feedback strategies for interactive learning tasks. In J. M. Spector, M. D. Merrill, J. J. G. van Merriënboer, & M. P. Driscoll (Eds.), *Handbook of research on educational communications and technology* (3rd ed., pp. 125–144). Lawrence Erlbaum.
- Narciss, S. (2012). Feedback in instructional contexts. In N. Seel (Ed.), *Encyclopedia of the sciences of learning* (Vol. F (6), pp. 1285–1289). Springer.
<https://doi.org/10.1007/978-1-4419-1428-6>
- Narciss, S. (2013). Designing and evaluating tutoring feedback strategies for digital learning. *Digital Education Review*, 23, 7–26.
- Narciss, S. (2017). Conditions and effects of feedback viewed through the lens of the interactive tutoring feedback model. In D. Carless, S. Bridges, C. Chan, & R. Glofcheski (Eds.), *Scaling up Assessment for Learning in Higher Education. The Enabling Power of Assessment* (Vol. 5, pp. 173–189). Springer.
https://doi.org/10.1007/978-981-10-3045-1_12
- Narciss, S., & Huth, K. (2004). How to design informative tutoring feedback for multimedia learning. In H. M. Niegemann, D. Leutner, & R. Brünken (Eds.), *Instructional Design for Multimedia Learning* (pp. 181–195). Waxmann.

References

- Narciss, S., Sosnovsky, S., Schnaubert, L., Andrès, E., Eichelmann, A., Gogvadze, G., & Melis, E. (2014). Exploring feedback and student characteristics relevant for personalizing feedback strategies. *Computers & Education*, 71, 56–76.
<https://doi.org/10.1016/j.compedu.2013.09.011>
- Nelson, M. E., Fiatarone, M. A., Morganti, C. M., Trice, I., Greenberg, R. A., & Evans, W. J. (1994). Effects of high-intensity strength training on multiple risk factors for osteoporotic fractures. A randomized controlled trial. *JAMA*, 272(24), 1909–1914.
<https://doi.org/10.1001/jama.1994.03520240037038>
- Nesi, H., & Gardner, S. (2012). Families of genres of assessed writing. In H. Nesi & S. Gardner (Eds), *Genres across the disciplines: Student writing in higher education* (pp. 21–56). Cambridge University Press. <https://doi.org/10.1017/9781009030199>
- Ngo, T. T.-N., Chen, H. H.-J., & Lai, K. K.-W. (2022). The effectiveness of automated writing evaluation in EFL/ESL writing: A three-level meta-analysis. *Interactive Learning Environments*, 1(1), 1–18. <https://doi.org/10.1080/10494820.2022.2096642>
- Nicol, D. J., & MacFarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199–218. <https://doi.org/10.1080/03075070600572090>
- Nietfeld, J. L., Cao, L., & Osborne, J. W. (2006). The effect of distributed monitoring exercises and feedback on performance, monitoring accuracy, and self-efficacy. *Metacognition Learning*, 1, 159–179. <https://doi.org/10.1007/s10409-006-9595-6>
- Nihalani, P. K., Mayrath, M., & Robinson, D. H. (2011). When feedback harms and collaboration helps in computer simulation environments: An expertise reversal effect. *Journal of Educational Psychology*, 103(4), 776–785.
<https://doi.org/10.1037/a0025276>
- Nunan, D., Aronson, J., & Bankhead, C. (2018). Catalogue of bias: Attrition bias. *BMJ evidence-based medicine*, 23(1), 21–22. <https://doi.org/10.1136/ebmed-2017-110883>

- Nunes, A., Cordeiro, C., Limpo, T., & Castro, S. L. (2021). Effectiveness of automated writing evaluation systems in school settings: A systematic review of studies from 2000 to 2020. *Journal of Computer Assisted Learning*, 38(2), 599–620.
<https://doi.org/10.1111/jcal.12635>
- OECD. (2023). *PISA 2022 Results (Volume I): The State of Learning and Equity in Education*. OECD Publishing. <https://doi.org/10.1787/53f23881-en>
- Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, 45(4), 867–872. <https://doi.org/10.1016/j.jesp.2009.03.009>
- O'Rourke, S. T., Calvo, R. A., & McNamara, D. S. (2011). Visualizing topic flow in students' essays. *Journal of Educational Technology & Society*, 14, 4–15.
- Ossner, J. (2006). Kompetenzen und Kompetenzmodelle im Deutschunterricht [Competencies and competency models in German lessons]. *Didaktik Deutsch* [German didactics], 11(21), 5–19.
- Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan – A web and mobile app for systematic reviews. *Systematic Reviews*, 5, 210.
<https://doi.org/10.1186/s13643-016-0384-4>
- Ozuru, Y., Best, R., Bell, C., Witherspoon, A., & McNamara, D. S. (2007). Influence of question format and text availability on the assessment of expository text comprehension. *Cognition and Instruction*, 25(4), 399–438.
<https://doi.org/10.1080/07370000701632371>
- Paas, F. (1992). Training strategies for attaining transfer of problem solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, 84(4), 429–434. <https://doi.org/10.1037/0022-0663.84.4.429>

References

- Paas, F., & van Gog, T. (2006). Optimising worked example instruction: Different ways to increase germane cognitive load. *Learning and Instruction, 16*(2), 87–91.
<https://doi.org/10.1016/j.learninstruc.2006.02.004>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ (Clinical research ed.), 372*(71).
<https://doi.org/10.1136/bmj.n71>
- *Palermo, C. J. (2017). *A framework for deliberate practice: Self-regulated strategy development and an automated writing evaluation program* [Doctoral dissertation, North Carolina State University]. North Carolina State University ProQuest Dissertations Publishing, 10610664. <https://www.proquest.com/dissertations-theses/framework-deliberate-practice-self-regulated/docview/1906983534/se-2?accountid=14433>
- Panadero, E., Andrade, H., & Brookhart, S. (2018). Fusing self-regulated learning and formative assessment: A roadmap of where we are, how we got here, and where we are going. *The Australian Educational Researcher, 45*(1), 13–31.
<https://doi.org/10.1007/s13384-018-0258-y>
- Panadero, E., & Lipnevich, A. A. (2022). A review of feedback models and typologies: Towards an integrative model of feedback elements. *Educational Research Review, 35*, 100416. <https://doi.org/10.1016/j.edurev.2021.100416>
- Patchan, M., Schunn, C., & Correnti, R. (2016). The nature of feedback: how peer feedback features affect students' implementation rate and quality of revisions. *Journal of Educational Psychology, 108*(8), 1098–1120. <https://doi.org/10.1037/edu0000103>

- Pekrun, R. (2005). Progress and open problems in educational emotion research. *Learning and Instruction*, 15, 497–506. <https://doi.org/10.1016/j.learninstruc.2005.07.014>
- Pekrun, R. (2017). Emotion and achievement during adolescence. *Child Development Perspectives*, 11(3), 215–221. <https://doi.org/10.1111/cdep.12237>
- Pekrun, R., & Bühner, M. (2014). Self-report measures of academic emotions. In R. Pekrun, & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 561–566). Routledge.
- Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist*, 37(2), 91–105.
https://doi.org/10.1207/S15326985EP3702_4
- Pham, B., Platt, R., McAuley, L., Klassen, T. P., & Moher, D. (2001). Is there a "best" way to detect and minimize publication bias? An empirical evaluation. *Evaluation & the Health Professions*, 24(2), 109–125.
<https://doi.org/10.1177/016327870102400202>
- Philipp, M. (2014). *Selbstreguliertes Schreiben. Schreibstrategien erfolgreich vermitteln [Self-regulated writing. Teaching writing strategies successfully]*. Beltz.
- Philipp, M. (2017). Wirksame Schreibförderung – Metaanalytische Befunde [Effective writing support - Meta-analytic findings]. In M. Becker-Mrotzek, J., Grabowski, & T. Steinhoff (Eds.), *Forschungshandbuch empirische Schreibdidaktik [Research manual for empirical writing didactics]* (pp. 187–202). Waxmann.
- Picard, R. W., Papert, S., Bender, W., Blumberg, B., Breazeal, C., Cavallo, D., Machover, T., Resnick, M., Roy, D., & Strohecker, C. (2004). Affective Learning — A Manifesto. *BT Technology Journal*, 22, 253–269.
<https://doi.org/10.1023/B:BTTJ.0000047603.37042.33>

References

- Pirnay-Dummer, P., & Ifenthaler, D. (2011). Reading guided by automated graphical representations: How model-based text visualizations facilitate learning in reading comprehension tasks. *Instructional Science*, *39*(6), 901–919.
<https://doi.org/10.1007/s11251-010-9153-2>
- Potter, R. F., & Bolls, P. (2011). *Psychophysiological measurement and meaning. Cognitive and emotional processing of media* (1st ed.). Routledge.
<https://doi.org/10.4324/9780203181027>
- Powell, P. (2009). Retention and writing instruction: Implications for access and pedagogy. *College Composition and Communication*, *60*, 664–682.
<https://www.jstor.org/stable/40593424>
- Raaijmakers, S. F., Baars, M., Schaap, L., Paas, F., & van Gog, T. (2017). Effects of performance feedback valence on perceptions of invested mental effort. *Learning and Instruction*, *51*, 36–46. <https://doi.org/10.1016/j.learninstruc.2016.12.002>
- Renkl, A. (1997). Learning from worked-out examples: A study on individual differences. *Cognitive Science*, *21*(1), 1–29. https://doi.org/10.1207/s15516709cog2101_1
- Renkl, A. (2005). The worked-out examples principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 229–245). Cambridge University Press. <https://doi.org/10.1017/CBO9780511816819.016>
- Renkl, A. (Ed.). (2008). *Lehrbuch Pädagogische Psychologie* [Manual of pedagogical psychology]. Huber.
- Renkl, A. (2009). Lehren und Lernen [Teaching and learning]. In R. Tippelt, & B. Schmidt-Hertha (Eds.), *Handbuch Bildungsforschung* [Handbook of educational research] (pp. 737–751). Verlag für Sozialwissenschaften.
- Renkl, A. (2013). Why practice recommendations are important in use-inspired basic research and why too much caution is dysfunctional. *Educational Psychology Review*, *25*(3), 317–324. <https://doi.org/10.1007/s10648-013-9236-0>

- 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>
- Rhöneck, C. von (1986). Vorstellungen vom elektrischen Stromkreis [Concepts of the electric circuit]. *Naturwissenschaften im Unterricht – Physik, Chemie* [Sciences in the classroom – physics, chemistry], 34(13), 10–14.
- Richter, J., Scheiter, K., & Eitel, A. (2016). Signaling text-picture relations in multimedia learning: A comprehensive meta-analysis. *Educational Research Review*, 17, 19–36. <https://doi.org/10.1016/j.edurev.2015.12.003>
- 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>
- Roscoe, R. D., Brandon, R. D., Snow, E. L., & McNamara, D. S. (2013). Game-based writing strategy practice with the Writing Pal. *Exploring technology for writing and writing instruction*, 1–20. <https://doi.org/10.4018/978-1-4666-4341-3.ch001>
- Roscoe, R. D., Jacovina, M. E., Allen, L. K., Johnson, A. C., & McNamara, D. S. (2016). Toward revision-sensitive feedback in automated writing evaluation. In T. Barnes, M. Chi., & M. Feng (Eds.), *Proceedings of the 9th International Conference on Educational Data Mining* (pp. 628–629). International Education Data Mining Society (IEDNS). https://www.educationaldatamining.org/EDM2016/proceedings/paper_80.pdf
- Roscoe, R. D., & McNamara, D. (2013). Writing Pal: Feasibility of an intelligent writing strategy tutor in the high school classroom. *Journal of Educational Psychology*, 105, 1010–1025. <https://doi.org/10.1037/a0032340>
- Roscoe, R. D., Snow, E. L., Allen, L. K., & McNamara, D. (2015). Automated detection of essay revising patterns: Applications for intelligent feedback in a writing tutor. *Technology, Instruction, Cognition and Learning*, 10, 59–79.

References

- *Roscoe, R. D., Snow, E. L., & McNamara, D. S. (2013). Feedback and revising in an intelligent tutoring system for writing strategies. In H. C. Lane, K. Yacef, J. Mostow, & P. Pavlik (Eds.), *Artificial Intelligence in Education. Lecture Notes in Computer Science* (Vol. 7926, pp. 259–268). Springer. https://doi.org/10.1007/978-3-642-39112-5_27
- Rosenshine, B. (2008). Five meanings of direct instruction. *Center on Innovation & Improvement*, 1–10.
- Rosnow, R. L., & Rosenthal, R. (1996). Contrasts and interactions redux: Five easy pieces. *Psychological Science*, 7(4), 253–257. <https://doi.org/10.1111/j.1467-9280.1996.tb00369>
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68–78. <https://doi.org/10.1037/0003-066X.55.1.68>
- Ryan, R. S., Wilde, M., & Crist, S. (2013). Compared to a small, supervised lab experiment, a large, unsupervised web-based experiment on a previously unknown effect has benefits that outweigh its potential costs. *Computers in Human Behavior*, 29(4), 1295–1301. <https://doi.org/10.1016/j.chb.2013.01.024>
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119–144. <https://doi.org/10.1007/bf00117714>
- Salden, R. J. C. M., Koedinger, K. R., Renkl, A., Aleven, V., & McLaren, B. M. (2010). Accounting for beneficial effects of worked examples in tutored problem solving. *Educational Psychology Review*, 22(4), 379–392. <https://doi.org/10.1007/s10648-010-9143-6>
- Scaife, M., & Rogers, Y. (1996). External cognition: How do graphical representations work? *International Journal of Human-Computer Studies*, 45(2), 185–213. <https://doi.org/10.1006/ijhc.1996.0048>

- Schardt, C., Adams, M. B., Owens, T., Keitz, S., & Fontelo, P. (2007). Utilization of the PICO framework to improve searching PubMed for clinical questions. *BMC Medical Informatics and Decision Making*, 7(16). <https://doi.org/10.1186/1472-6947-7-16>
- Schecker, H., Wilhelm, T., Hopf, M., & Duit, R. (2018). *Schülervorstellungen und Physikunterricht: Ein Lehrbuch für Studium, Referendariat und Unterrichtspraxis* [Student perceptions and physics education: A textbook for studying, teacher training, and teaching practice.]. Springer. <https://doi.org/10.1007/978-3-662-57270-2>
- Scherer, R., Siddiq, F., & Teo, T. (2015). Becoming more specific: Measuring and modeling teachers' perceived usefulness of ICT in the context of teaching and learning. *Computers & Education*, 88, 202–214. <https://doi.org/10.1016/j.compedu.2015.05.005>
- Scheurer, J., Campos, J. A., Chan, J. S., Chen, A., Cho, K., & Perez, E. (2022). *Training language models with natural language feedback*. arXiv:2204.14146. <https://doi.org/10.48550/arXiv.2204.14146>
- Schipolowski, S., & Böhme, K. (2016). Assessment of writing ability in secondary education: Comparison of analytic and holistic scoring systems for use in large-scale assessments. *L1-Educational Studies in Language and Literature*, 16, 1–22.
- Schmidt, S. J. (2003). Was wir vom Lernen zu wissen glauben. In Deutsches Institut für Erwachsenenbildung (DIE), E. Nuisl, C. Schiersmann, & H. Siebert (Eds.), *Literatur- und Forschungsreport Weiterbildung – Gehirn und Lernen* (26. Jahrgang 2003, Heft 3/2003, pp. 40–50). W. Bertelsmann Verlag GmbH & Co. KG. https://ams-forschungsnetzwerk.at/downloadpub/gehirn_und_lernen.pdf#page=40
- Schneider, J., Backfisch, I., & Lachner, A. (2021). Facilitating open science practices for research syntheses: PreregRS guides preregistration. *Research Synthesis Methods*, 13(2), 284–289. <https://doi.org/10.1002/jrsm.1540>

References

- Schneider, S., Wirzberger, M., & Rey, G. D. (2019). The moderating role of arousal on the seductive detail effect in a multimedia learning setting. *Applied Cognitive Psychology*, 33(1), 71–84. <https://doi.org/10.1002/acp.3473>
- Schwarzer, G., Carpenter, J. R., & Rücker, G. (2015). *Meta-analysis with R*. Springer. <https://doi.org/10.1007/978-3-319-21416-0>
- Scriven, M. (1967). The Methodology of Evaluation. In R. W. Tyler, R. M. Gagne, & M. Scriven (Eds.), *Perspectives of Curriculum Evaluation Chicago* (pp. 39–83). Rand McNally.
- Sharp, L. A. (2016). Preservice teachers and writing: An analysis of academic performance - RESEARCH. *Kentucky Journal of Excellence in College Teaching and Learning*, 13(2). <https://encompass.eku.edu/kjectl/vol13/iss1/2>
- Shermis, M. D., & Burstein, J. (Eds.). (2003). *Automated essay scoring: A cross-disciplinary perspective*. Lawrence Erlbaum Associates Publishers.
- Shermis, M. D., Burstein, J., & Bliss, L. (2004, April). *The impact of automated essay scoring on high stakes writing assessments* [Paper presentation]. The annual meetings of the National Council on Measurement in Education, San Diego, CA.
- Shin, J., Guo, Q., & Gierl, M. J. (2019). Multiple-choice item distractor development using topic modeling approaches. *Frontiers in Psychology*, 10, 825. <https://doi.org/10.3389/fpsyg.2019.00825>
- Shirah, J. F., & Sidney, P. G. (2023). Computer-based feedback matters when relevant prior knowledge is not activated. *Learning and Instruction*, 87, 101796. <https://doi.org/10.1016/j.learninstruc.2023.101796>
- Shute, V. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153–189. <https://doi.org/10.3102/0034654307313795>
- Sibley, L., Lachner, A., Plicht, C., Fabian, A., Backfisch, I., Scheiter, K., & Bohl, T. (2023). *Feasibility of adaptive teaching with technology: Which implementation*

- conditions matter?* [Manuscript submitted for publication]. Institute of Education, University of Tübingen.
- Sinha, T., & Kapur, M. (2021). When problem solving followed by instruction works: Evidence for productive failure. *Review of Educational Research, 91*(5), 761–798. <https://doi.org/10.3102/00346543211019105>
- Sterne, J. A. C., Gavaghan, D., & Egger, M. (2000). Publication and related bias in meta-analysis. *Journal of Clinical Epidemiology, 53*(11), 1119–1129. [https://doi.org/10.1016/S0895-4356\(00\)00242-0](https://doi.org/10.1016/S0895-4356(00)00242-0)
- Stevenson, M., & Phakiti, A. (2014). The effects of computer-generated feedback on the quality of writing. *Assessing Writing, 19*, 51–65. <https://doi.org/10.1016/j.asw.2013.11.007>
- Strobl, C., Ailhaud, E., Benetos, K., Devitt, A., Kruse, O., Proske, A., & Rapp, C. (2019). Digital support for academic writing: A review of technologies and pedagogies. *Computers and Education, 131*, 33–48. <https://doi.org/10.1016/j.compedu.2018.12.005>
- Sturm, A. (2014). Experten- und Novizen-Feedback in der Domäne Schreiben [Expert and novice feedback in the domain of writing]. *Leseforum.ch, 3*, 1–22.
- Sturm, A. (2016). Beurteilen und Kommentieren von Texten als fachdidaktisches Wissen [Evaluating and commenting on texts as pedagogical content knowledge]. *Leseräume, 3*(3), 115–132.
- Swart, E. K., Nielen, T. M. J., & Sikkema-de Jong, M. T. (2019). Supporting learning from text: A meta-analysis on the timing and content of effective feedback. *Educational Research Review, 28*, 100296. <https://doi.org/10.1016/j.edurev.2019.100296>
- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*(4), 295–312. [https://doi.org/10.1016/0959-4752\(94\)90003-](https://doi.org/10.1016/0959-4752(94)90003-)

References

- Sweller, J. (2004). Instructional design consequences of an analogy between evolution by natural selection and human cognitive architecture. *Instructional Science*, 32, 9–31 (2004). <https://doi.org/10.1023/B:TRUC.0000021808.72598.4d>
- Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive Load Theory*. Springer. <https://doi.org/10.1007/978-1-4419-8126-4>
- Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10, 251–296.
- Tan, A. L., Gillies, R., & Jamaludin, A. A. (2021). A case study: Using a neuro-physiological measure to monitor students' interest and learning during a Micro:Bit activity. *Education Sciences*, 11(8), 379. <https://doi.org/10.3390/educsci11080379>
- Tsai, C.-T., Chen, J.-J., Yang, C.-Y., & Chang, J. S. (2020). LinggleWrite: A coaching system for essay writing. In A. Celikyilmaz, & T.-H. Wen (Eds.), *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations* (pp. 127–133). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.acl-demos.17>
- Urban-Woldron, H., & Hopf, M. (2012). Entwicklung eines Testinstruments zum Verständnis in der Elektrizitätslehre [Development of a diagnostic instrument for testing student understanding of basic electricity concepts]. *Zeitschrift für Didaktik der Naturwissenschaften* [Journal for didactics of sciences], 18, 201–227.
- van Berkum, J. J. A., & de Jong, T. (1991). Instructional environments for simulations. *Education and Computing*, 6(3-4), 305–358. [https://doi.org/10.1016/0167-9287\(91\)80006-J](https://doi.org/10.1016/0167-9287(91)80006-J)
- Vancouver, J. B., & Morrison, E. W. (1995). Feedback inquiry: The effect of source attributes and individual differences. *Organizational Behavior and Human Decision Processes*, 62(3), 276–285. <https://doi.org/10.1006/obhd.1995.1050>

- van der Kleij, F. M., Eggen, T. J., Timmers, C. F., & Veldkamp, B. P. (2012). Effects of feedback in a computer-based assessment for learning. *Computers & Education*, *58*(1), 263–272. <https://doi.org/10.1016/j.compedu.2011.07.020>
- van der Kleij, F. M., Feskens, R. C. W., & Eggen, T. J. H. M. (2015). Effects of feedback in a computer-based learning environment on students' learning outcomes: A meta-analysis. *Review of Educational Research*, *85*(4), 475–511. <https://doi.org/10.3102/0034654314564881>
- van Gog, T., Kester, L., & Paas, F. (2011). Effects of worked examples, example-problem, and problem-example pairs on novices' learning. *Contemporary Educational Psychology*, *36*(3), 212–218. <https://doi.org/10.1016/j.cedpsych.2010.10.004>
- van Gog, T., & Paas, F. (2008). Instructional efficiency: Revisiting the original construct in educational research. *Educational Psychologist*, *43*(1), 16–26. <http://dx.doi.org/10.1080/00461520701756248>
- van Gog, T., Paas, F., & van Merriënboer, J. J. (2006). Effects of principle-oriented worked examples on trouble shooting transfer performance. *Learning and Instruction*, *16*(2), 154–164. <https://doi.org/10.1016/j.learninstruc.2006.02.003>
- van Gog, T., Paas, F., & van Merriënboer, J. J. (2008). Effects of studying sequences of principle-oriented and procedure-oriented worked examples on troubleshooting transfer efficiency. *Learning and Instruction*, *18*(3), 211–222. <https://doi.org/10.1016/j.learninstruc.2007.03.003>
- van Gog, T., & Rummel, N. (2010). Example-based learning: Integrating cognitive and social-cognitive research perspectives. *Educational Psychology Review*, *22*(2), 155–174. <https://doi.org/10.1007/s10648-010-9134-7>
- van Harsel, M., Hoogerheide, V., Verkoeijen, P., & van Gog, T. (2019). Effects of different sequences of examples and problems on motivation and learning.

References

- Contemporary Educational Psychology*, 58, 260–275.
<https://doi.org/10.1016/j.cedpsych.2019.03.005>
- van Harsel, M., Hoogerheide, V., Verkoeijen, P., & van Gog, T. (2020). Examples, practice problems, or both? Effects on motivation and learning in shorter and longer sequences. *Applied Cognitive Psychology*, 34(4), 793–812.
<https://doi.org/10.1002/acp.3649>
- VanLehn, K. (2008). *Mind bugs: The origins of procedural misconceptions*. MIT Press.
- van Merriënboer, J. J., Clark, R. E., & De Croock, M. (2002). Blueprints for complex learning: The 4C/ID-model. *Educational technology research and development*, 50(2), 39–61. <https://doi.org/10.1007/BF02504993>
- van Steendam, E., Rijlaarsdam, G., van den Bergh, H., & Sercu, L. (2014). The mediating effect of instruction on pair composition in L2 revision and writing. *Instructional Science*, 42(6), 905–927. <https://www.jstor.org/stable/43575485>
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1–48. <https://doi.org/10.18637/jss.v036.i03>
- Villalon, J., & Calvo, R. A. (2011). Concept maps as cognitive visualizations of writing assignments. *Journal of Educational Technology & Society*, 14(3), 16–27.
- Wade-Stein, D., & Kintsch, E. (2004). Summary Street: Interactive computer support for writing. *Cognition and Instruction*, 3, 333–362.
https://doi.org/10.1207/s1532690xci2203_3
- 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>
- *Wang, E. L., Matsumura, L. C., Correnti, R., Litman, D., Zhang, H., Howe, E., Magooda, A., & Quintana, R. (2020). eRevis(ing): Students’ revision of text evidence use in an

- automated writing evaluation system. *Assessing Writing*, 44, 100449.
<https://doi.org/10.1016/j.asw.2020.100449>
- Weinstein, C., & Mayer, R. (1986). The Teaching of Learning Strategies. In M. Wittrock (Ed.), *Handbook of Research on Teaching* (pp. 315–327). Macmillan.
- *Weston-Sementelli, J. L., Allen, L. K., & McNamara, D. S. (2016). Comprehension and Writing Strategy Training Improves Performance on Content-Specific Source-Based Writing Tasks. *International Journal of Artificial Intelligence in Education*, 1–32.
<https://doi.org/10.1007/s40593-016-0127-7>
- Wharton, C., & Kintsch, W. (1991). An overview of the construction-integration model: A theory of comprehension as a foundation for a new cognitive architecture. *SIGART Bulletin*, 2(4), 169–173. <https://doi.org/10.1145/122344.122379>
- 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>
- William, D. (2007). Keeping learning on track: Classroom assessment and the regulation of learning. In F. K. Lester Jr. (Ed.), *Second handbook of mathematics teaching and learning* (pp. 1053–1098). Information Age Publishing.
- Williams, S. E. (1997, March). Teachers' written comments and students' responses: A socially constructed interaction. *Proceedings of the annual meeting of the Conference on College Composition and Communication*, Phoenix, AZ.
- Wilson, J., & Roscoe, R. D. (2020). Automated writing evaluation and feedback: Multiple metrics of efficacy. *Journal of Educational Computing Research*, 58, 87–125.
<https://doi.org/10.1177/0735633119830764>
- Wilson, J., Roscoe, R., & Ahmed, Y. (2017). Automated formative writing assessment using a levels of language framework. *Assessing Writing*, 34, 16-36.
<http://doi.org/10.1016/j.asw.2017.08.002>

References

- Winne, P. H., & Butler, D. (1994). Student Cognition in learning from teaching. In T. Husen & T. Postlewaite (Eds.), *International Encyclopedia of Education* (2nd ed., pp. 5738–5745). Pergamon.
- Wischgoll, A. (2017). Improving undergraduates' and postgraduates' academic writing skills with strategy training and feedback. *Frontiers in Education*, 2(33), 1–15. <https://doi.org/10.3389/feduc.2017.00033>
- Wise, S. L., Hake, B. S., Pozehl, B. J., Barnes, L. B., & Lukin, L. E. (1989). Providing item feedback in computer-based tests: Effects of initial success and failure. *Educational and Psychological Measurement*, 49, 479–486.
- Wisniewski, B., Zierer, K., & Hattie, J. (2020). The Power of Feedback Revisited: A Meta-Analysis of Educational Feedback Research, *Frontiers in Psychology*, 10(3087). <https://doi.org/10.3389/fpsyg.2019.03087>
- Wittrock, M. C. (1974). Learning as a generative process. *Educational Psychologist*, 11(2), 87–95. <https://doi.org/10.1080/00461527409529129>
- Wittrock, M. C. (1990). Generative Processes of Comprehension. *Educational Psychologist*, 24, 345–376. https://doi.org/10.1207/s15326985ep2404_2
- Wittwer, J., & Renkl, A. (2008). Why instructional explanations often do not work: A framework for understanding the effectiveness of instructional explanations. *Educational Psychologist*, 43(1), 49–64. <https://doi.org/10.1080/00461520701756420>
- Wu, Y., Henriksson, A., Nouri, J., Duneld, M., & Li, X. (2022). Beyond benchmarks: Spotting key topical sentences while improving automated essay scoring performance with topic-aware BERT. *Electronics*, 12(1), 150. <https://doi.org/10.3390/electronics12010150>
- Yang, C., Luo, L., Vadiello, M. A., Yu, R., & Shanks, D. R. (2021). Testing (quizzing) boosts classroom learning: A systematic and meta-analytic review. *Psychological Bulletin*, 147(4), 1–37. <http://dx.doi.org/10.1037/bul00003091>

- Yeung, K. L., Carpenter, S. K., & Corral, D. (2021). A comprehensive review of educational technology on objective learning outcomes in academic contexts. *Educational Psychology Review*, 33, 1583–1630. <https://doi.org/10.1007/s10648-020-09592-4>
- Zhai, N., & Ma, X. (2022). The effectiveness of automated writing evaluation on writing quality: A meta-analysis. *Journal of Educational Computing Research*, 61(4). <https://doi.org/10.1177/07356331221127300>
- *Zhu, M., Liu, O. L., & Lee, H.-S. (2020). The effect of automated feedback on revision behavior and learning gains in formative assessment of scientific argument writing. *Computers & Education*, 143, 103668. <https://doi.org/10.1016/j.compedu.2019.103668>
- *Zhu, M., Lee, H.-S., Wang, T., Liu, O. L., Belur, V., & Pallant, A. (2017). Investigating the impact of automated feedback on students' scientific argumentation. *International Journal of Science Education*, 39(12), 1648–1668. <https://doi.org/10.1080/09500693.2017.1347303>
- Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). Academic press. <https://doi.org/10.1016/B978-012109890-2/50031-7>

List of Figures

Figure 1. <i>MISCA model by Panadero and Lipnevich (2022)</i>	14
Figure 2. <i>The CAFCALI Model</i>	33
Figure 3. <i>Assignment of the Studies into the CAFCALI Model</i>	43
Figure 4. <i>Kognitive und metakognitive Schreibprozesse, adaptiert nach Hayes (2012)</i>	50
Figure 5. <i>Beispiel einer in CohViz generierten Concept Map zum Thema Osmose</i>	65
Figure 6. <i>Procedure According to the PRISMA Framework (Inspired by Page et al., 2021)</i>	83
Figure 7. <i>Excerpt of the Strategy Instruction (Translated)</i>	127
Figure 8. <i>Excerpt of the Control Instruction for the No-Instruction Condition (Translated)</i>	130
Figure 9. <i>Example of Feedback for Practice Task 1 (Translated)</i>	132
Figure 10. <i>Moderated Mediation Analysis Regarding the Impact of the Valence of Feedback and Students' Arousal</i>	144
Figure 11. <i>Example of Negative Feedback of Practice Task 1 in Experiment 2, if the Answer was Incorrect (Translated)</i>	147
Figure 12. <i>Revised CAFCALI Model</i>	178

List of Tables

Table 1. <i>Computerbasierte Feedbacksysteme im Vergleich hinsichtlich der postulierten Hauptdimensionen</i>	59
Table 2. <i>Moderator Variables</i>	85
Table 3. <i>Overview of all Included Studies in the Meta-Analyses</i>	91
Table 4. <i>Overview of Effect Sizes and Confidence Intervals of all Included Studies in the Meta-Analysis of Studies with Pretest-Posttest Design (k = 24)</i>	95
Table 5. <i>Effect Sizes and Confidence Intervals of all Included Studies in the Meta-Analysis of Studies with Pretest-Posttest-Control Group-Design (k = 15)</i>	97
Table 6. <i>Results of Moderator Analysis with Studies with Pretest-Posttest Design</i>	100
Table 7. <i>Results of Moderator Analysis with Studies with Pretest-Posttest-Control Group Design</i>	103
Table 8. <i>Means and Standard Deviations for all Measurements in Experiment 1</i>	137
Table 9. <i>Results of the Multilevel Analysis Exploring the Effects of Feedback, Strategy Instruction, and its Interaction on Students' Practice Tasks, Overall, Near, and far Transfer Performance in Experiment 1</i>	140
Table 10. <i>Means and Standard Deviations for all Measurements in Experiment 2</i>	149
Table 11. <i>Results of the Multilevel Analysis Exploring the Effects of Feedback, Strategy Instruction, and its Interaction on Students' Practice Tasks, Overall, Near, and far Transfer Performance in Experiment 2</i>	152
Table 12. <i>Means and Standard Deviations for all Measurements in Experiment 3</i>	159
Table 13. <i>Results of the Multilevel Contrast Analysis in Experiment 3</i>	162

Appendix

Appendix A. Correlations of the Variables Across Conditions in Experiment 1

Variable	1	2	3	4	5	6	7	8	9	10	11	12
1. Prior knowledge												
2. Performance in practice tasks	(.23* .34											
	.24* .41*)											
3. Performance in transfer tasks	(.24* .32*	(.66* .60*										
	.39* .23*)	.68* .52*)										
4. Performance in near transfer tasks	(.20* .28*	(.57* .55*	(.87 .84*									
	.29* .31*)	.70* .53*)	.84* .82*)									
5. Performance in far transfer tasks	(.21 .23*	(.55* .43*	(.83* .78*	(.43* .32*								
	.36* .09)	.40* .35*)	.80* .85*)	.34* .41*)								
6. Subjective difficulty after learning phase	(-.22* -.27*	(-.34* -.44*	(-.36* -.33*	(-.36* -.35*	(-.26* -.17							
	-.34* -.36*)	-.39* -.39*)	-.47* -.27*)	-.39* -.29*)	-.37* -.17)							
7. Mental effort after learning phase	(-.30* -.05	(-.28* .18	(-.17 .25*	(-.15 .19*	(-.13 .21	(.40* .08						
	-.12 -.30*)	.21* -.09)	.20* -.00)	.18 -.03)	.15 .03)	.17 .21*)						
8. Arousal after learning phase	(-.03 -.16	(.08 .23*	(.10 .10	(.06 .09	(.10 .08	(.09 -.20*	(.13 .07					
	.04 .05)	.18 .17)	.03 .21*)	.07 .15)	-.02 .20*)	-.14 -.20*)	.30* -.02)					
9. Subjective difficulty after practice phase	(-.42* -.38*	(-.55* -.72*	(-.50* -.52*	(-.41* -.51*	(-.45* -.32*	(-.40* .54*	(.38* -.10	(-.13 -.18				
	-.43* -.42*)	-.59* -.77*)	-.62* -.50*)	-.58* -.49*)	-.43* -.36*)	.64* .58*)	.04 .18)	-.13 -.31*)				

Note. * indicates $p < .05$.

Appendix A. Correlations of the Variables Across Conditions in Experiment 1, *continued*

Variable	1	2	3	4	5	6	7	8	9	10	11	12
1. Prior knowledge												
2. Performance in practice tasks												
3. Performance in transfer tasks												
4. Performance in near transfer tasks												
5. Performance in far transfer tasks												
6. Subjective difficulty after learning phase												
7. Mental effort after learning phase												
8. Arousal after learning phase												
9. Subjective difficulty after practice phase												
10. Mental effort after practice phase	(-.27* -.29* -.32* -.31*)	(-.21* -.08 -.07 -.35*)	(-.14 .01 -.07 -.16)	(-.12 -.04 -.00 -.21*)	(-.13 .07 -.12 -.07)	(.16 .17 .29* .48*)	(.43* .44* .64* .56*)	(-.01 .19* .24 -.07)	(.55* .38* .39* .54*)			
11. Arousal after practice phase	(-.04 -.11 .06 .02)	(.10 .41* .30* .22*)	(.13 .21* .15 .25*)	(.11 .19* .14 .18)	(.10 .15 .10 .23*)	(.12 -.21* -.18 -.22*)	(.14 .16 .27* .04)	(.87* .85* .88* .89*)	(-.21* -.31* -.25* -.36*)	(-.01 .22* .18 -.02)		
12. Monitoring accuracy practice tasks	(-.05 -.14 -.20* -.20*)	(-.78* -.79* -.45* -.78*)	(-.42* -.46* -.73* -.27*)	(-.32* -.38* -.58* -.31*)	(-.40* -.37* -.61* -.15)	(.13 .08 .09 .14)	(.05 -.35* -.27* -.09)	(-.04 -.14 -.08 -.01)	(.18 .44* .29* .47*)	(.03 -.05 -.06 .07)	(-.03 -.34* -.22* -.10)	
13. Monitoring accuracy transfer tasks	(.01 -.06 -.20* .03)	(-.33* -.32* -.45* -.23*)	(-.62* -.71* -.73* -.66*)	(-.49* -.55* -.58* -.53*)	(-.57* -.61* -.61* -.58*)	(.06 -.10 .09* -.05)	(-.15 -.41* -.26* -.21*)	(-.04 .00 .08 -.01)	(-.01 .16 .23* .13)	(-.11 -.15 -.09 -.19)	(-.03 -.12 -.06 -.09)	(.65* .66* .75* .53*)

Note. * indicates $p < .05$.

Appendix B. Summary of ANCOVAs Measuring the Effects of Feedback, Strategy Instruction, and its Interaction on Students' Practice Tasks, Overall, Near, and far Transfer Performance in Experiment 1

Variables	$F(1, 432)$	p	η_p^2
Practice tasks			
Strategy Instruction	9.21	.003	.02
Feedback	0.34	.562	.00
Interaction	2.15	.144	.01
Transfer tasks (overall)			
Strategy Instruction	4.19	.041	.01
Feedback	2.07	.151	.01
Interaction	0.36	.551	.00
Near transfer			
Strategy Instruction	0.51	.474	.00
Feedback	2.20	.139	.01
Interaction	1.34	.247	.00
Far transfer			
Strategy Instruction	7.19	.008	.02
Feedback	0.72	.397	.00
Interaction	4.98	.026	.01

Note. Significant p -values are highlighted in bold.

Appendix C. Summary of Mediation Analyses Regarding the Instruction-by-Feedback Interaction Effect on Students' Far Transfer in Experiment 1

Mediator	<i>b</i>	<i>SE_b</i>	95% CI
Mental effort	.0003	.0014	[-0.0029, 0.0032]
Subjective difficulty	.0033	.0067	[-0.0097, 0.0166]
Judgement of learning	-.0005	.0060	[-0.0125, 0.0110]
Monitoring accuracy	.0186	.0099	[-0.0008, 0.0382]
Arousal	-.0008	.0025	[-0.0060, 0.0043]
Pleasure	-.0009	.0033	[-0.0077, 0.0054]

Note. Significant results are highlighted in bold letters.

Appendix D. Summary of Moderated Mediation with Valence of Feedback as Moderator in Experiment 1

Mediator	<i>b</i>	<i>SE_b</i>	95% CI
Mental effort	-.0068	.0058	[-0.0206, 0.0019]
Subjective difficulty	-.0003	.0116	[-0.0232, 0.0225]
Judgement of learning	.0051	.0113	[-0.0172, 0.0270]
Monitoring accuracy	.0335	.0262	[-0.0168, 0.0859]
Arousal	.0136	.0076	[0.0009, 0.0303]
Pleasure	.0071	.0083	[-0.0088, 0.0247]

Note. Significant results are highlighted in bold letters.

Appendix E. Correlations of the Variables in Experiment 2

Variable	1	2	3	4	5	6
1. Prior knowledge						
2. Performance in practice tasks	(.27* .25* .26* .32*)					
3. Performance in transfer tasks	(.31* .26* .20 .14)	(.47* .40* .68* .51*)				
4. Performance in near transfer tasks	(.34* .26* .22 .19)	(.52* .37* .76* .61*)	(.89* .81* .84* .81*)			
5. Performance in far transfer tasks	(.19 .16 .12 .05)	(.28* .27* .38* .27*)	(.85* .78* .83* .86*)	(.53* .27* .39* .41*)		
6. Subjective difficulty after learning phase	(-.17 -.19 -.35* -.33*)	(-.25* -.30* -.41* -.31*)	(-.34* -.36* -.32* -.37*)	(-.33* -.30* -.23* -.36*)	(-.25* -.26* -.30* -.27*)	
7. Mental effort after learning phase	(-.22 -.32* .09 -.29*)	(-.08 -.19 .18 .01)	(-.19 -.21 .07 -.04)	(-.07 -.22 .16 .00)	(-.28* -.11 -.05 -.06)	(.20 .24* .15 .32*)
8. Arousal after learning phase	(.25* -.03 .19 .11)	(.28* .12 .16 .12)	(.16 -.06 .22 .21)	(.16 -.04 .26* .15)	(.10 -.06 .11 .20)	(-.37* -.12 -.24* -.14)
9. Subjective difficulty after practice phase	(-.32* -.16 -.20 -.27*)	(-.37* -.65* -.47* -.79*)	(-.46* -.35* -.41* -.52*)	(-.41* -.42* -.39* -.59*)	(-.40* -.13 -.29* -.30*)	(.40* .34* .53* .47*)
10. Mental effort after practice phase	(-.17 -.10 .04 -.40*)	(-.17 -.26* -.10 -.25*)	(-.12 -.11 -.16 -.12)	(.02 -.15 -.09 -.11)	(-.25* -.03 -.18 -.09)	(.02 .21 .25* .33*)
11. Arousal after practice phase	(.23* .04 .13 .08)	(.37* .29* .12 .23*)	(.24* -.00 .26* .25*)	(.24* .08 .28* .19)	(.17 -.09 .14 .23*)	(-.44* -.02 -.14 -.12)
12. Monitoring accuracy practice tasks	(-.21 -.27* .05 -.11)	(-.88* -.81* -.74* -.70*)	(-.33* -.28* -.45* -.20)	(-.38* -.28* -.59* -.26*)	(-.19 -.16 -.16 -.09)	(.05 .02 -.15 -.16)
13. Monitoring accuracy transfer tasks	(-.09 -.14 .19 .16)	(-.23* .27* -.17 .14)	(-.71* -.56* -.54* -.49*)	(-.63* -.42* -.37* -.25*)	(-.62* -.47* -.53* -.54*)	(.06 -.05 -.25* -.13)
14. Utility of the learning phase	(-.04 -.10 .14 -.01)	(.06 -.03 .44* .14)	(-.02 .05 .47* .25*)	(.01 -.00 .47* .16)	(-.05 .08 .32* .26*)	(-.25* -.18 -.53* -.21)
15. Utility of the practice phase	(.27* .05 .21 -.06)	(.38* .43* .51* .46*)	(.34* .31* .33* .40*)	(.37* .34* .39* .42*)	(.20 .15 .16 .27*)	(-.20 -.17 -.50* -.05)

Note. * indicates $p < .05$.

Appendix E. Correlations of the Variables in Experiment 2, *continued*

Variable	7	8	9	10	11	12	13	14
1. Prior knowledge								
2. Performance in practice tasks								
3. Performance in transfer tasks								
4. Performance in near transfer tasks								
5. Performance in far transfer tasks								
6. Subjective difficulty after learning phase								
7. Mental effort after learning phase								
8. Arousal after learning phase	(.02 .03 .19 .15)							
9. Subjective difficulty after practice phase	(.21 .40* .18 .10)	(-.33* -.22 -.14 -.14)						
10. Mental effort after practice phase	(.46* .53* .56* .40*)	(-.07 .02 .06 .10)	(.37* .56* .54* .36*)					
11. Arousal after practice phase	(-.07 -.17 .17 .19)	(.86* .73* .91* .88*)	(-.37* -.30* -.17 -.22)	(-.03 -.18 .05 .11)				
12. Monitoring accuracy practice tasks	(-.03 .19 -.20 -.08)	(-.07 .13 .08 .12)	(.11 .41* .12 .40*)	(.16 .22 -.01 .13)	(-.12 -.15 .04 .00)			
13. Monitoring accuracy transfer tasks	(-.01 .02 -.05 .02)	(.12 .35* .06 .13)	(-.15 -.34* -.34* -.21)	(-.03 -.12 -.11 -.02)	(.09 .32* -.00 .15)	(.37* .04 .34* .23*)		
14. Utility of the learning phase	(.27* .25* .20 .29*)	(.33* .26* .44* .20)	(-.04 .21 -.32* -.23)	(.15 .18 -.02 .11)	(.27* .21 .40* .24*)	(.02 .09 -.08 .04)	(.00 -.10 -.07 -.00)	
15. Utility of the practice phase	(-.19 -.22 .20 .33*)	(.15 .32* .25* .13)	(-.57* -.40* -.48* -.45*)	(-.07 -.19 .00 .15)	(.32* .36* .22 .23*)	(-.17 -.33* -.23* -.25*)	(.13 .12 .23* .10)	(-.01 .26* .59* .51*)

Note. * indicates $p < .05$.

Appendix F. Summary of ANCOVAs Measuring the Effects of Feedback, Strategy Instruction, and its Interaction on Students' Practice Tasks, Overall, Near, and far Transfer Performance in Experiment 2

Variables	$F(1, 305)$	p	η_p^2
Practice Tasks			
Strategy Instruction	3.78	.051	.01
Feedback	1.93	.166	.00
Interaction	0.37	.545	.00
Transfer tasks (overall)			
Strategy Instruction	6.43	.012	.02
Feedback	8.39	.004	.02
Interaction	0.81	.369	.00
Near transfer			
Strategy Instruction	0.58	.447	.00
Feedback	10.18	.002	.02
Interaction	0.19	.662	.00
Far transfer			
Strategy Instruction	12.02	.001	.04
Feedback	2.67	.103	.01
Interaction	1.13	.289	.00

Note. Significant p -values are highlighted in bold.

Appendix G. Correlations of the Variables in Experiment 3

Variable	1	2	3	4	5	6
1. Prior knowledge						
2. Performance in practice tasks	(.41* .23 .32*)					
3. Performance in transfer tasks	(.11 .32* .27)	(.42* .40* .47*)				
4. Performance in near transfer tasks	(.34* .34* .32*)	(.45* .41* .42*)	(.85* .85* .86*)			
5. Performance in far transfer tasks	(-.22 .17 .10)	(.21 .22 .35*)	(.77* .76* .80*)	(.33* .31* .38*)		
4. Subjective difficulty after learning phase	(-.16 -.22 -.21)	(-.17 -.36* -.20)	(-.16 -.45* -.30*)	(-.23 -.44* -.30*)	(-.01 -.28* -.19)	
5. Mental effort after learning phase	(-.18 .10 -.03)	(-.03 .20 .10)	(.08 -.03 .11)	(-.05 -.01 .16)	(.20 -.04 .02)	(.30* .18 -.12)
6. Arousal after learning phase	(-.13 .19 .07)	(.01 .41* .33*)	(.11 .29* .31*)	(.12 .27* .24)	(.05 .19 .29*)	(.05 -.27* -.31*)
7. Subjective difficulty after practice phase	(-.27* -.32* -.37*)	(-.40* -.80* -.71*)	(-.29* -.50* -.44*)	(-.25 -.46* -.41*)	(-.21 -.35* -.32*)	(.45* .60* .25)
8. Mental effort after practice phase	(-.04 -.13 .05)	(-.04 -.24 -.24)	(-.08 -.29* -.09)	(-.13 -.21 .01)	(.00 -.26* -.18)	(.36* .38* .05)
9. Arousal after practice phase	(-.15 .28* .11)	(.19 .47* .30*)	(.03 .26* .17)	(.09 .21 .12)	(-.06 .21 .17)	(-.14 -.21 -.10)
10. Monitoring accuracy after learning phase	(-.35* -.01 -.07)	(-.83* -.76* -.76*)	(-.26* -.16 -.20)	(-.24 -.18 -.16)	(-.18 -.07 -.17)	(-.11 -.12 -.06)
11. Monitoring accuracy after practice phase	(-.00 .04 .08)	(-.08 .26* .04)	(-.70* -.52* -.63*)	(-.54* -.41* -.47*)	(-.62* -.44* -.58*)	(-.19 -.20 .12)
12. Utility of the learning phase	(.12 .13 .11)	(.04 .45* .47*)	(.31* .50* .53*)	(.32* .45* .50*)	(.17 .34* .38*)	(-.45* -.37* -.48*)
13. Utility of the practice phase	(.12 .13 .11)	(.30* .56* .51*)	(.19 .33* .42*)	(.13 .40* .38*)	(.18 .10 .30*)	(-.37* -.30* -.35*)

Note. * indicates $p < .05$.

Appendix G. Correlations of the Variables in Experiment 3, *continued*

Variable	7	8	9	10	11	12	13	14
1. Prior knowledge								
2. Performance in practice tasks								
3. Performance in transfer tasks								
4. Performance in near transfer tasks								
5. Performance in far transfer tasks								
4. Subjective difficulty after learning phase								
5. Mental effort after learning phase								
6. Arousal after learning phase	(.00 .25* .43*)							
7. Subjective difficulty after practice phase	(-.04 -.06 .03)	(-.06 -.28* -.22)						
8. Mental effort after practice phase	(.34* .43* .44*)	(.29* .01 .16)	(.18 .53* .22)					
9. Arousal after practice phase	(.07 .09 .28)	(.81* .80* .82*)	(-.12 -.31* -.22)	(.23 -.07 .30*)				
10. Monitoring accuracy after learning phase	(.15 -.22 .11)	(.11 -.15 -.10)	(.06 .39* .50*)	(.05 .03 .36*)	(.00 -.29* -.13)			
11. Monitoring accuracy after practice phase	(-.10 .12 -.01)	(-.02 .12 -.13)	(-.24 -.32* -.03)	(-.15 -.15 .08)	(.12 .10 .01)	(.23 .05 .21)		
12. Utility of the learning phase	(-.00 .23 .28)	(.12 .28* .51*)	(-.27 -.45* -.38*)	(-.18 -.07 .14)	(.17 .34* .41*)	(.17 -.26* -.11)	(-.07 -.09 -.23)	
13. Utility of the practice phase	(-.05 .26 .15)	(.19 .35* .28)	(-.59* -.47* -.43*)	(.01 -.01 .11)	(.18 .44* .31*)	(-.04 -.47* -.27)	(.22 .03 -.04)	(.45* .68* .58*)

Note. * indicates $p < .05$.

Appendix H. Results of the Contrast Analyses in Experiment 3

Variables	<i>t</i> (162)	<i>p</i>	<i>d</i>
Practice tasks			
Control vs. experimental condition (H1)	2.33	.021	0.34
Flipped vs. regular condition (H2)	0.55	.580	0.07
Transfer tasks (overall)			
Control vs. experimental condition (H1)	2.34	.027	0.37
Flipped vs. regular condition (H2)	0.21	.831	0.03
Near transfer			
Control vs. experimental condition (H1)	2.12	.036	0.33
Flipped vs. regular condition (H2)	-0.36	.716	-0.08
Far transfer			
Control vs. experimental condition (H1)	1.52	.131	0.25
Flipped vs. regular condition (H2)	0.80	.427	0.15

Note. Significant *p*-values are highlighted in bold. For H1, we used the following contrast weights: control condition: -2; regular sequence: 1;

flipped sequence: 1. For H2, we used the following contrast weights: control condition: 0; regular sequence: -1; flipped sequence: 1.

Furthermore, we included students' prior knowledge as covariate in both contrast analyses.

Declaration on Contributions

Although the publication-based dissertation is written as a monography, it includes contents of three manuscripts which were written together with other authors. The proportional contributions to the manuscripts are presented in the subsequent tables.

Chapter 2 Feedback – Ja, klar?!

Author	Author position	Scientific ideas in %	Data generation in %	Analysis & interpretation in %	Paper writing in %
Salome Wagner	First	60	-	80	70
Andreas Lachner	Second	40	-	20	30

Title of paper: Feedback – Ja, klar?! Digitale Medien zur Förderung von Schreibkompetenzen

Publication process: Published: *leseforum.ch*, 3

ISSN: 2624-7771

Chapter 3 Does Computer-Based Feedback Foster Expository Writing?

Author	Author position	Scientific ideas in %	Data generation in %	Analysis & interpretation in %	Paper writing in %
Salome Wagner	First	70	80	70	80
Jürgen Schneider	Second	5	20	20	0
Iris Backfisch	Third	0	0	0	10
Katharina Scheiter	Fourth	10	0	0	0
Andreas Lachner	Fifth	15	0	10	10

Title of paper: Does Computer-Based Feedback Foster Expository Writing? A Meta-Analysis

Publication process: Invited for major revision: Educational Psychology Review

Chapter 4 The more, the better?

Author	Author position	Scientific ideas in %	Data generation in %	Analysis & interpretation in %	Paper writing in %
Salome Wagner	First	65	90	80	65
Leonie Sibley	Second	0	0	10	5
David Weiler	Third	0	5	0	5
Jan-Philipp Burde	Fourth	0	5	0	5
Katharina Scheiter	Fifth	10	0	0	5
Andreas Lachner	Sixth	25	0	10	15

Title of paper: The more, the better? Learning with feedback and instruction

Publication process: Published: *Learning and Instruction*, 89, 101844
<https://doi.org/10.1016/j.learninstruc.2023.101844>