Automated Formative Feedback for Enhancing Summarizing Skills:

The Impact of Feedback Characteristics, Learner Factors, and Instructional Prompts on Feedback Engagement and Effectiveness in Higher Education

Vom Promotionsausschuss des Fachbereichs Erziehungswissenschaften der Rheinland-Pfälzischen Technischen Universität Kaiserslautern-Landau zur Verleihung des akademischen Grades Doktor der Philosophie (Dr. phil.) genehmigte Dissertation.

Dissertation

vorgelegt am 13.12.2024

von Veronika Barkela

Datum der wissenschaftlichen Aussprache: 25.04.2025

Vorsitz des Promotionsausschusses: Prof. Dr. Pascal Bastian, RPTU Kaiserslautern-Landau

Erstgutachterin: Prof. Dr. Miriam Leuchter, RPTU Kaiserslautern-Landau

Zweitgutachter: Prof. Dr. Henrik Saalbach, Universität Leipzig

Contents

List of Figures ... III

List of Tables ... III

List of Abbreviations ... III

Acknowledgements ... IV

Abstract ... V

1. General Introduction ... 1
    - 1.1 The Interactive Tutoring Feedback Model ... 5
    - 1.2 Characteristics of Automated Formative Feedback ... 7
    - 1.3 Learner Factors ... 12
        - 1.3.1 Engagement with Automated Formative Feedback ... 12
        - 1.3.2 Motivation ... 18
        - 1.3.3 Cognitive Load ... 23
    - 1.4 Instructional Factors ... 25
2. The Present Dissertation ... 30
    - 2.1 Effectiveness of Automated Formative Feedback ... 37
    - 2.2 Impact of Cognitive and Motivational Resources on Feedback Engagement ... 39
    - 2.3 Interplay of Instructional Prompts and Automated Formative Feedback ... 40
3. General Discussion ... 42
    - 3.1 Discussion of the Research Questions ... 42
        - 3.1.1 Research Question 1.1 ... 42
        - 3.1.2 Research Question 1.2 ... 46
        - 3.1.3 Research Question 2.1 ... 48
        - 3.1.4 Research Question 2.2 ... 50
        - 3.1.5 Research Question 3.1 ... 54
        - 3.1.6 Research Question 3.2 ... 58
    - 3.2 Overall Contribution of the Dissertation to the Field ... 62
    - 3.3 Critical Evaluation and Limitations of the Studies ... 65
    - 3.4 Outlook on Future Research ... 70
    - 3.5 Conclusion ... 74

References 76
Appendix 105
Manuscript 1 107
Manuscript 2 141
Manuscript 3 177
Curriculum Vitae 217
Declaration of Originality 222

List of Figures

Figure 1 Interactive Tutoring Feedback Model (Narciss, 2006, 2013) 7

Figure 2 Simplified ITF Model Applied to Summarizing Tasks (Narciss, 2006, 2013) 44

List of Tables

Table 1 Comparison of the Feedback Engagement and Feedback Literacy Constructs 15

Table 2 Indicators of Cog. Feedback Engagement in Feedback Processing and Revision 16

Table 3 Overview of the Three Presented Manuscripts.. 36

Table 4. Summary of the Key Findings of the Three Presented Manuscripts 61

List of Abbreviations

EFL English as a Foreign Language

ITF Interactive Tutoring Feedback

LLM Large Language Model

LSA Latent Semantic Analysis

Acknowledgements

I would like to express my heartfelt gratitude to my supervisor, Prof. Dr. Miriam Leuchter, for her adventurous spirit in embarking on this project with me and for her unwavering support in the realization of the three publications presented in this dissertation. I am deeply grateful for her guidance, which has allowed me to pursue my research interests and develop as a scientist. I am also sincerely thankful to Prof. Dr. Henrik Saalbach for his immediate willingness to take on the task of writing the second review.

My gratitude extends to my esteemed colleagues at the Institute of Child and Youth Education at the RPTU. Thank you for the enriching discussions and lunch times full of laughter. In particular, I want to express my heartfelt thanks to my dear colleagues and friends, Heide Sasse and Dr. Sabrina Stiel-Dämmer. Your time, encouragement, ideas, and steadfast support have been invaluable to me. It is a pleasure to work alongside you, and your friendship truly enriches my life.

I am deeply grateful to my family, without whom I could not have reached this milestone. Thank you, Berend, for your constant support—you make life easier and so much more interesting. To my parents, I am endlessly thankful for instilling in me a belief in myself and encouraging me to reach for the stars. Moreover, I wish to express my gratitude to my parents and parents-in-law for stepping in whenever time was scarce. Lastly, my deepest thanks go to my children, Tamme and Lotte, for their understanding, support, and patience during this journey. You are the joy of my life and my greatest pride!

Abstract

Formative feedback is widely recognized for its effectiveness in enhancing learning processes, such as improving summarizing skills. However, delivering individualized formative feedback remains a significant challenge for university educators, particularly in large courses. Automated formative feedback systems offer a promising solution by providing timely, actionable support at scale. However, a thorough understanding of both the provider's and receiver's automated formative feedback processes is needed when effectively integrating such systems into university teaching practices.

Therefore, this dissertation investigates whether and under which conditions automated formative feedback effectively fosters summarizing skills among university students, addressing three key dimensions with three studies: feedback characteristics, individual learner factors, and instructional factors. To investigate these dimensions in depth, the first study examined the effectiveness of automated formative feedback in fostering summarizing skills. The findings reveal that automated formative feedback significantly improves summarizing skills, especially when students engage in multiple feedback loops. A second study investigated the joint influences of motivation, mental effort, and feedback engagement over time, revealing their significant impact on learners' engagement with feedback. The findings underscore the need to foster students' willingness to allocate mental resources to seek feedback more frequently. With a third study, instructional prompts were compared to automated formative feedback and their combined effects were examined. While prompts supported summarizing skills, they were less effective than feedback, and their combination did not yield additional benefits, indicating potential substitution effects. The studies trace the intricate processes involved in providing and utilizing automated formative feedback, while shedding light on its strengths and unrealized potential. These findings and their implications for learners and instructors in higher education will be discussed.

1. General Introduction

Formative feedback effectively enhances learning processes (Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Shute, 2008). However, providing formative feedback regularly has long been a challenge for university teachers, particularly those managing large courses, due to the significant time commitment required. In a world where digital learning technologies are fundamentally altering the ways teachers teach and students learn, automated formative feedback offers solutions. Computers have evolved from being passive tools to active pedagogical agents, guiding learners through their educational journeys (Kabudi et al., 2021; Maier & Klotz, 2022; Martin et al., 2020; Shi & Aryadoust, 2024). Advances in areas such as natural language processing have expanded the potential for applying high pedagogical standards and offering valuable support, including formative feedback, through digital technologies (cf. Strobl et al., 2019). This progress provides an opportunity to offer students the individualized support they need to learn effectively without placing an excessive burden on teachers. Moreover, it enables the creation of additional learning opportunities that foster not only subject-specific knowledge but also overarching competencies necessary for academic success, such as concise writing and effective summarizing. However, effectively integrating such systems into university teaching practices requires a comprehensive understanding of automated formative feedback processes involved in providing and receiving automated formative feedback. Therefore, this dissertation investigates if and under which conditions automated formative feedback effectively fosters summarizing skills among university students.

Summarizing is an important strategy for university students as they are expected to quickly extract, assimilate and concisely articulate relevant information from the scientific literature (Kürschner & Schnotz, 2007; van Dijk & Kintsch, 1983). To this end, summarizing has been shown to promote active engagement with course material and to encourage learners to distill complex concepts into concise, manageable forms (Dole et al., 1991; Mok & Chan, 2016; Rinehart et al., 1986). Additionally, summarizing enhances comprehension and retention, as it requires learners to process, synthesize, and integrate information into their cognitive frameworks (Dunlosky et al., 2013; Kirkland & Saunders, 1991; McAnulty, 1981). Summarizing involves multiple cognitive processes, such as activating prior knowledge, selecting relevant information, aligning new insights with pre-existing knowledge, integrating new information into the cognitive schema by constructing a mental model of the text, and translating this understanding into coherent written text (Becker-Mrotzek et al., 2014; Hidi & Anderson, 1986; Kellogg & Raulerson, 2007; Perin et al., 2017; Wade-Stein & Kintsch, 2004; Westby et al., 2010). The quality of a summary depends on a learner's mental model of the original text (M. K. Kim & McCarthy, 2021; Schnotz, 2006). Mental models are representations of a text that include both the explicitly stated information as well as inferences made by linking related details with prior knowledge (van Dijk & Kintsch, 1983). Individuals with more prior knowledge and better coordination of these processes are more capable of constructing comprehensive mental models and producing higher-quality summaries (Hathorn & Rawson, 2012; Kellogg, 1987; K. Kim et al., 2019; Y.-S. G. Kim, 2020; Mason et al., 2013; McCarthy & McNamara, 2021).

Consequently, summarizing has the potential to foster reading comprehension (Head et al., 1989; Hill, 1991; E. Kintsch, 1990; E. Kintsch et al., 2000; Lenhard et al., 2012; Leopold et al., 2019) and it may also strengthen text composition skills (J. Li, 2014; Schoonen, 2019). By condensing essential information and communicating it clearly and succinctly, students may develop the skill to produce well-structured, coherent, non-redundant, and precise texts. Furthermore, the dual benefit of fostering reading comprehension and text composition skills makes summarizing an effective learning strategy, as it enables learners to better organize and articulate complex ideas (Dunlosky et al., 2013; Stevens et al., 2019). However, undergraduates often struggle to coordinate the cognitive processes of summarizing and thus to grasp the gist of a scientific text, discern core ideas from peripheral information, and link new information to their prior knowledge (Kintsch, 1990; Duke & Pearson, 2009). As a result, they face difficulties in producing concise summaries in their own words thus needing support in developing effective summarizing skills (Keck, 2006; M. K. Kim & McCarthy, 2021).

Feedback provides this support by setting clear criteria for summary quality and guiding students toward meeting these criteria. Specifically, such criteria include incorporating core information, using one's own words, minimizing redundancies, and ensuring the summary is significantly shorter than the original text (Lenhard et al., 2012; Sung et al., 2016; Wade-Stein & Kintsch, 2004). Research has already shown that individualized formative feedback can foster reading and writing skills as means for summarizing (Graham, 2018; Schunk & Rice, 1991). Feedback that aligns with the cognitive processes involved in summarizing can help students adjust their internal representations of a high-quality summary by regularly comparing their work to an external standard. However, providing such support for an overarching competence is not always feasible for university teachers, particularly in large classes and with limited resources (Allen et al., 2016; Perin, 2019; Shanahan, 2019). An automated feedback system may offer a solution to this challenge (cf. Deeva et al., 2021). Several successful feedback systems have been introduced for example to foster the development of academic writing and the processing of scientific texts (M. K. Kim & McCarthy, 2021; Proske et al., 2012; also cf. Strobl et al., 2019).

To investigate if and under which conditions automated formative feedback effectively fosters summarizing skills among university students, factors of the system and of the learners need to be considered. Narciss (2013), therefore, proposes a threefold approach to design and evaluate automated formative feedback, including the characteristics of feedback (1), individual learner factors (2), and instructional factors (3). In all three areas, significant research gaps exist for the example of fostering summarizing skills of university students with automated formative feedback. 1) With regard to the characteristics of the feedback, research has focused so far on improving reading comprehension in elementary school students (Lenhard et al., 2012; Sung et al., 2016; Wade-Stein & Kintsch, 2004). Studies that investigate summarizing skills in university students particularly fostering these with automated formative feedback over several time points, however, are scarce (Chew et al., 2019; M. K. Kim & McCarthy, 2021). 2) Although several variables concerning learner prerequisites on the feedback process have already been investigated (Han, 2019; Seifried et al., 2016), the joint influences of learner's feedback engagement and individual factors such as motivational and cognitive resources have not yet been thoroughly explored. 3) With regard to instructional factors, especially prompts present themselves as a promising alternative or complement to feedback for effectively communicating with students and scaffolding their learning processes. Several studies have shown that prompts targeting summarizing strategies can effectively enhance summarizing skills (Ahn, 2022; Doolittle et al., 2006; Green & Holman, 2021; Palinscar & Brown, 1984; Tolosa et al., 2015). However, the support potential of prompts compared to automated formative feedback has not been systematically investigated. Moreover, the interaction and complementary effects of instructional prompts and automated formative feedback remain unclear, presenting a significant gap in the current research that warrants further exploration.

These research gaps are in the focus of this dissertation. Building on the Interactive Tutoring Feedback (ITF) Model (Narciss, 2006, 2013, 2017), the present work investigates automated formative feedback from three perspectives: the feedback message itself, learner factors, and instructional factors. To address these objectives, the dissertation first presents and analyzes the theoretical background, existing findings, and research gaps in detail (chapters 1.1 – 1.4). From this theoretical analysis, specific research questions are derived (chapter 2). These questions were investigated in three manuscripts that are included in the appendix and summarized in chapters 2.1 – 2.3. Chapter 3 provides a synthesis of the findings, discusses their implications and limitations, and outlines perspectives for future research.

1.1 The Interactive Tutoring Feedback Model

Feedback has consistently attracted the attention of researchers and practitioners in offline (Bangert-Drowns et al., 2004; Evans, 2013; Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Lipnevich & Panadero, 2021; Shute, 2008; Wisniewski et al., 2020) and online learning environments (Maier & Klotz, 2022; Martin et al., 2020; Shi & Aryadoust, 2024). Feedback refers to information provided by an agent (e.g., teacher, peer, self, computer) regarding a learner's performance or understanding, with the goal of guiding future improvement and bridging the gap between current and desired outcomes (Hattie & Timperley, 2007; Narciss, 2006, 2013; Shute, 2008). Numerous studies and meta-analyses have demonstrated that feedback can be highly effective in enhancing learning processes (Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Shute, 2008). However, many of the studies so far rely on feedback models that primarily focus on the conditions and design of feedback. This implies that these models only implicitly account for the communicative and interactive nature of feedback processes. However, even the best-designed learning environments and feedback strategies are only effective if students actively engage with them. Newer models, particularly those that are concerned with computer-generated feedback, therefore not only consider the feedback itself but also cognitive and motivational processes on the part of the receiver, as well as the dynamic interaction between provider and receiver (Lipnevich & Panadero, 2021). In this context, the ITF model (Narciss, 2006, 2013, 2017) is a promising approach.

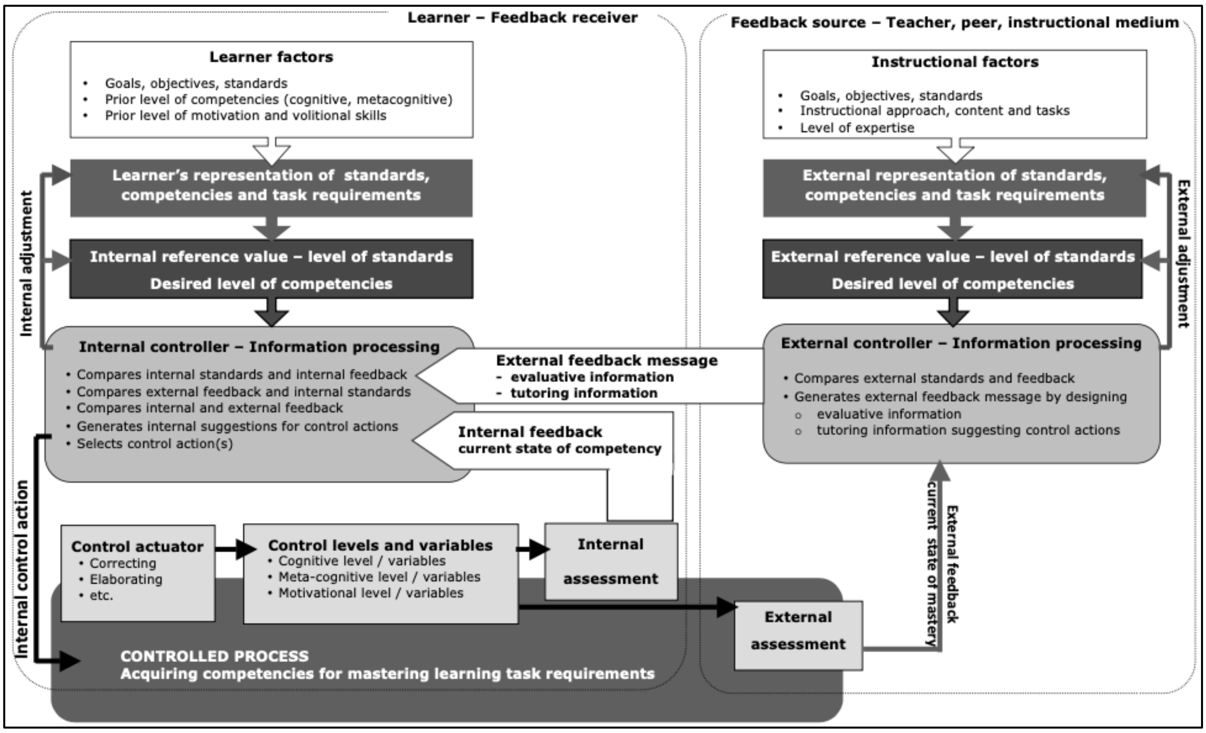
The ITF model distinguishes between internal and external feedback loops and relates them to learning outcomes (Figure 1). Internal feedback represents the endpoint of a learner's internal evaluation of a task, where they assess whether the task has been completed according to their internal reference standards. When, for example, drafting a summary, learners utilize these internal reference standards, which are shaped by their personal interpretations of external standards. Furthermore, the internal reference standards are also influenced by the learners' understanding of the task requirements, metacognitive abilities, and motivation. Through the process of comparing their work to these internal standards, learners generate internal feedback, which informs their decision to either revise their summary or conclude the task.

In contrast, external feedback is provided by an external agent who defines external reference standards based on instructional factors and representations of standards, competencies, and task requirements. In this dissertation, the external standard for a summary includes communicating all relevant information, using own words, and being concise and short which is reflected in an expert summary. The feedback then involves comparing the learner's summary against this external standard and typically includes individualized feedback messages that may identify errors or suggest strategies for improvement. Upon receiving external feedback, learners interpret this information through their personal cognitive, metacognitive, and motivational frameworks, aligning it with their internal reference standards and competencies. As part of this process, learners generate internal feedback by comparing the external standards and feedback to their own work and pre-existing internal standards. Successful learners use these comparisons to identify necessary control actions and choose appropriate activities to enhance their learning outcomes.

The interplay between internal and external feedback highlights the complexity of designing and evaluating feedback systems. A key challenge lies in bridging the two feedback loops to not only provide accurate and actionable information but also consider learners' individual prerequisites, and providers' instructional objectives. To address this challenge, 1) characteristics of the feedback, 2) learners' individual factors, and 3) instructional factors have to be examined (Narciss, 2013, 2017).

Figure 1

Interactive Tutoring Feedback Model (Narciss, 2006, 2013)



1.2 Characteristics of Automated Formative Feedback

Feedback is characterized by its a) function, b) content, and c) presentation. a) The function of feedback can be cognitive, metacognitive, or motivational, aiming to enhance understanding, promote self-regulation, or increase engagement, respectively (Narciss et al., 2014). For example, in the context of fostering summarizing, automated formative feedback following a cognitive function can provide explicit hints which words were copied from the original text, or which sentences contain irrelevant information. b) The content of feedback is highly context-dependent, and may range from simple knowledge of results to more elaborate, specific, and informative guidance that helps learners understand not just what is wrong but how to improve (Black & Wiliam, 2009). Its effectiveness is significantly influenced by where it directs the learner's attention. Feedback that focuses on the task or specific task strategies, relates to goals, or offers cues or reinforcements tends to enhance performance, as it helps learners identify errors and correct them effectively (Kluger & DeNisi, 1996; Sadler, 2006). For example, providing semantic feedback that marks problematic sentences according to assessment criteria specifies where and how revising is necessary. c) The presentation of feedback can vary significantly, being either summative or formative, either adaptive to the learner's needs or fixed, and delivered immediately or delayed (Shute, 2008). Research consistently highlights the critical role of formative feedback, especially when delivered promptly and in iterative cycles. This timely and repeated feedback helps learners connect the two feedback loops, making it more relevant, actionable, and effective in guiding their subsequent efforts (Hattie & Timperley, 2007; Kluger & DeNisi, 1996). Additionally, formative feedback supports learners in monitoring and regulating their own learning processes by offering actionable insights (Nicol & Macfarlane-Dick, 2006). Furthermore, it fosters improvement over time by encouraging continuous revision and refinement (van der Kleij et al., 2015).

This is where automated formative feedback excels, meeting the demand for timeliness and consistency, even for large numbers of students (Seifried et al., 2016; Wilson et al., 2014). Automated systems can deliver immediate feedback, allowing learners to quickly identify and correct errors or misconceptions (van Lehn, 2011), hereby reinforcing learning through real-time adjustments and iterative loops (Deeva et al., 2021; van der Kleij et al., 2015). Additionally, computer-generated feedback is accessible anytime and anywhere, giving learners the flexibility to review and act on feedback at their own pace.

Beyond accessibility and timeliness, automated feedback systems are designed to adapt to individual learners' needs and performance levels. By leveraging algorithms and data analytics, these systems track learning patterns over time, thus offering tailored feedback that addresses specific challenges and learning trajectories. Furthermore, this data supports educators by providing insights into learner progress, enabling more informed decisions about instructional strategies and targeted support (Meghji et al., 2020; Schmid et al., 2022; Seifried et al., 2015, 2017).

However, while computer-generated feedback is superior to analog feedback in providing immediate, reliable feedback (e.g., copied words), it faces a key challenge due to its reliance on algorithms that often lack contextual understanding. This limitation can result in less nuanced feedback and difficulties in addressing complex, higher-order tasks that require interpretation, judgment, or personalized dialogue (Lipnevich & Smith, 2009a). The quality of automated feedback is thus heavily dependent on the design and sophistication of the underlying algorithms.

On a sidenote, the introduction of large language models (LLM), such as ChatGPT (OpenAI, 2022), represents a significant advancement in the development of automated complex feedback. These models can provide more nuanced, context-aware feedback and personalized guidance by understanding complex language patterns and interpreting open-ended responses, mimicking a human-like dialogue (Shi & Aryadoust, 2024). Additionally, they can handle a wider range of topics, offer real-time explanations, examples, and clarifications, and adapt their responses based on the learner's input, creating a more interactive and tailored feedback experience. Despite their advantages, LLMs also face challenges as feedback providers, particularly in research contexts. For example, their feedback can often be overly generic, failing to address task-specific nuances or provide actionable suggestions, regardless of the model used (Banihashem et al., 2024). Additionally, responses from LLMs tend to vary depending on context and input, and they rarely build on previously provided feedback, leading to inconsistencies that undermine reliability (Chang et al., 2024). Furthermore, LLMs lack transparency in explaining their feedback. While they can generate responses based on probabilistic patterns, they often fail to provide clear reasoning behind their suggestions, making it difficult to assess their validity (Bender et al., 2021). Due to these challenges and the fact that the project underpinning this dissertation and its studies commenced before the release of publicly accessible LLMs, the studies in this dissertation utilized a feedback algorithm based on latent semantic analysis (LSA; Landauer, 1999). Consequently, the theoretical framework presented here primarily focuses on such approaches, while future perspectives on how this dissertation's findings align with recent innovations will be addressed in the general discussion.

Automated feedback for promoting summarizing skills can be designed to support and enhance the coordination of key cognitive processes while fostering revision (Wade-Stein & Kintsch, 2004). Information about content coverage ensures having correctly identified and understood the key concepts necessary to build an accurate mental representation of the text (Garner, 1982; Head et al., 1989; van Dijk & Kintsch, 1983). Moreover, information on relevance clarifies how effectively students identify and prioritize the essential aspects of a text, distinguishing key points from supplementary details (H. Li et al., 2018; Rivard, 2001). Providing feedback on relevance helps students refine their ability to filter out non-essential information and focus on the core message, a critical skill for accurate summarizing. Furthermore, feedback on the avoidance of copying words or passages from the text and expressing ideas in one's own words highlights the importance of processing new information and translating it into personal understanding (W. Kintsch & van Dijk, 1978; Mok & Chan, 2016). This encourages students to engage actively with the material, promoting deeper comprehension and the integration of new knowledge into existing cognitive frameworks. Additionally, feedback on redundancy avoidance focuses on helping students streamline the core message of the text, reorganize its ideas, and clearly articulate their mental models without unnecessary repetition (Hill, 1991; E. Kintsch, 1990). This guidance helps students refine their writing to ensure that their summaries are not only clear and concise but also effective in communicating their understanding of the material. Finally, feedback on adequate length of the summary reinforces the need for students to condense and reorganize information, which supports the creation of a coherent and efficient mental model (Dunlosky et al., 2013; Friend, 2001).

Feedback systems employing LSA for feedback on these textual features (in the following referenced to as task criteria) have shown to foster text comprehension as well as summarizing skills. For example, Wade-Stein and Kintsch (2004) developed a system which significantly improved the content quality of summaries written by elementary and middle school students. Lenhard et al. (2012) adapted this system for German elementary and middle school students, finding positive effects on reading comprehension and fluency. Sung et al. (2016) demonstrated that both semantic and concept feedback effectively enhanced summary content coverage for Chinese elementary students, with reduced tool reliance over time indicating improved summarizing skills. Therefore, adapting such a feedback system to the university context may also support university students in developing effective summarizing skills and thus, improve their reading comprehension and text composition (Graham, 2006;

Hill, 1991; E. Kintsch, 1990). Yet, it is unclear whether such a system supports all aspects of summarizing to the same extent and whether university students will effectively make use of the formative feedback in this context (Q. Guo et al., 2022; Zhai & Ma, 2021).

1.3 Learner Factors

Learners' individual factors significantly influence how they engage with and benefit from automated formative feedback, as these factors can either enhance or limit their ability to improve competencies toward desired standards. Motivation and cognitive load are particularly important in this context. Kinsey (2022) highlighted that individual engagement is strongly influenced by both learners' motivation and the cognitive load they experience. Therefore, this section first reviews the existing research on automated formative feedback engagement, followed by an examination of the theoretical foundations of motivation and cognitive load.

1.3.1 Engagement with Automated Formative Feedback

From a constructivist perspective, engagement is fundamental to learning, as it enables learners to actively build and expand their knowledge through the interaction with tasks, materials, or support resources (Evans, 2013; Malecka et al., 2022; Zimbardi et al., 2017). It encompasses the ongoing effort, emotional regulation, and active participation that learners invest to achieve their performance and learning goals (Fredricks et al., 2004; Jung & Lee, 2018; Sun & Rueda, 2012; S. Zhang & Liu, 2019). Educational research has long focused on engagement to explain various student behaviors, such as student alienation and dropout rates in schools (Fredricks et al., 2004), learning achievement and success (Jung & Lee, 2018; Wu & Kang, 2021), and the uptake of learning support (Sun & Rueda, 2012) and feedback (Ellis, 2010; Han & Hyland, 2015; To, 2022).

In particular, engagement with feedback has gained increasing attention since automated feedback systems have been advancing (Ali et al., 2018; Y. Liu & Sun, 2021; M. Liu et al., 2017; Hyland & Zhang, 2018; also cf. Pearson, 2024). In these systems, the communicative ability between receiver and provider is limited to the algorithm's level of sophistication. Therefore, the efficacy of automated feedback is greatly contingent upon how individuals engage with it (Handley et al., 2011; Price et al., 2011). Due to this imbalance, there has been a shift in investigating feedback processes from the perspective of the provider to the perspective of the receiver focusing on individuals' skills and capacities (Boud & Molloy, 2013; Nieminen & Carless, 2023). Specifically, two research trends in investigating student-centered feedback processes have emerged in recent years. Anchored in the framework of academic motivation (Eccles, 2016; Fredricks et al., 2004), the feedback engagement construct has lately been developed particularly by researchers in the field of English as a foreign language (Han, 2019; Han & Hyland, 2015; Y. Zhang & Gao, 2024; Z. Zhang & Hyland, 2022; Y. Zheng & Yu, 2018). Meanwhile, the concept of feedback literacy has been gaining widespread attention in higher education research (Carless & Boud, 2018; Carless & Winstone, 2023; Malecka et al., 2022; Nieminen & Carless, 2023; Sutton, 2012; Winstone et al., 2019).

Feedback engagement refers to the active participation, receptiveness, and responsiveness of individuals in the process of receiving and acting upon feedback (Handley et al., 2011; Pearson, 2024; Price et al., 2011). It encompasses the degree to which feedback recipients invest their attention, energy, and cognitive resources in understanding, reflecting upon, and applying feedback received from others. Therefore, feedback engagement is conceptualized as a three-dimensional construct which comprises cognitive, affective and behavioral dimensions (Ellis, 2010; Han & Hyland, 2015; Handley et al., 2011). Cognitive feedback engagement describes the extent of students' cognitive response and intensity of processing feedback to improve learning outcomes (Han, 2017; Price et al., 2011; Y. Zhang & Gao, 2024). Affective feedback engagement is associated with emotional and attitudinal reactions towards the feedback (Koltovskaia & Mahapatra, 2022; Z. Zhang & Hyland, 2022) or the perceived valence of the feedback (Mayordomo et al., 2022; Seifried et al., 2016). While behavioral feedback engagement refers to the activities initiated by the recipient after receiving feedback (Han & Xu, 2021; Y. Liu & Sun, 2021).

Similarly, according to Sutton (2012), feedback literacy is defined as a set of learning practices with which learners are empowered to appraise, evaluate, and interpret information within disciplines and assess the quality of their own and others' work. Again, the concept of feedback literacy comprises three dimensions: The epistemological dimension (knowing) includes the knowledge of the learning potential of feedback and how to use feedback to expand knowledge and improve in a certain domain. A feedback literate person uses feedback both summatively to assess skills and formatively to improve performance. This dimension refers to cognitive processes of decoding feedback and including insights to the learning process. The ontological dimension (being) involves the understanding how feedback affects identity and emotions. A feedback literate person is receptive to feedback while maintaining control over her emotions during the process. She is reflective and self-critical yet possesses the discernment to accurately assess her abilities within a specific domain. This dimension refers to affective processes of emotion control, reflectivity, and students' development of educational identities, e.g. how they view themselves as learners and engage with support like feedback. The practical dimension (acting) includes understanding the necessity of actively applying feedback to improve performance. A feedback literate person knows how to decode and interpret feedback as intended by the provider and translate it into specific actions. This dimension draws on processes such as reading, interpreting, and applying feedback to enhance learning outcomes (Chong, 2021; Winstone et al., 2019).

While feedback literacy can be seen as a more proactive and preparatory approach, focusing on equipping students with tools and understanding before they receive feedback, the feedback engagement construct takes a more reactive approach, emphasizing how students respond after receiving feedback (Molloy et al., 2020; Winstone et al., 2017). Yet, in comparison, both concepts show several similarities in their conceptualizations. Both aim to model the effective use of feedback from a learner-centered approach that acknowledges the tripartite structure of feedback reception. They recognize the importance of managing emotional responses to feedback and emphasize the need to take action based on feedback (Table 1). Importantly, both concepts understand feedback literacy, aka engagement with feedback, as a skill set which can be developed, measured, and traced (Nieminen & Carless, 2023; To, 2022). However, the operationalization of the three dimensions are highly discussed and differ depending on subjects, activities, and research designs (Pearson, 2024). To be consistent with the terminology of the second manuscript of this dissertation, the terminology of the feedback engagement construct will be used in the following sections.

Table 1
Comparison of the Feedback Engagement and Feedback Literacy Constructs

		Feedback engagement	**Feedback literacy**
Dimensions	cognitive response and intensity of processing feedback	cognitive	epistemological (knowing)
	emotional and attitudinal reactions, perceived valence, receptiveness	affective	ontological (being)
	actions initiated by receiver	behavioral	practical (acting)

Several studies attribute cognitive feedback engagement to cognitive and metacognitive operations of either processing feedback or the revision process. In their influential empirical study, Han and Hyland (2015) developed a framework for investigating these operations, utilizing the taxonomy of cognitive and metacognitive strategies established by Oxford (2011). Following their line of investigation several case studies have been conducted using written documents, revision operations, reflection reports, and interviews to elicit cognitive processes when learners receive written corrective feedback (Cheng & Zhang, 2024; Mayordomo et al., 2022; Tian & Zhou, 2020; Y. Zhang & Gao, 2024; Z. Zhang, 2020; Y. Zheng & Yu, 2018). They extracted several indicators for different levels of cognitive feedback engagement (Table 2).

Table 2
Indicators of Cognitive Feedback Engagement in Feedback Processing and Revision Processes

Feedback processing	Revision processes
Decision not to use feedback	No response to feedback
Reading the feedback	Corrections of grammar and mechanics
Noticing aspects of the feedback	Addition
Understanding the feedback	Deletion
Identifying aspects of the feedback in the work	Substitution
Applying feedback to improve task performance	Transformation

Furthermore, they identified indicators for metacognitive operations such as strategies for regulating mental processes, practices, and emotional reactions (Han & Hyland, 2015; Z. Zhang & Hyland, 2022). However, these studies revealed that cognitive and metacognitive operations often overlap and occur simultaneously, making it difficult to determine whether the use of these strategies was deliberate. Therefore, Y. Zhang and Gao (2024) argue that cognitive and metacognitive operations are conceptually and practically intertwined when revising a text due to feedback, and thus cannot be distinguished. Another approach aimed at analyzing the depth of feedback processing. Researchers focused on students' awareness of the feedback distinguishing between noticing and understanding (Han & Hyland, 2015; C. Lee, 2020; Y. Zheng & Yu, 2018) or assessed how well students reported they had understood the received feedback and made revisions accordingly (Yu, Zhou, et al., 2019). In conclusion, all presented approaches to measure cognitive feedback engagement attribute high cognitive feedback engagement to students who, due to the received feedback, perform operations that contribute to the improvement of their texts.

For affective feedback engagement, researchers aim to capture either students' affective and attitudinal responses to feedback (Han, 2017; To, 2022; Z. Zhang & Hyland, 2022) or the valence they attribute to the feedback (Mayordomo et al., 2022). In some studies, students were interviewed and asked for their emotional reactions such as anxiety, unpleasantness, and upset, when having received feedback (Han & Hyland, 2015; Koltovskaia & Mahapatra, 2022; Z. Zhang & Hyland, 2022). Furthermore, some studies investigated students' capacity to handle their emotions triggered by feedback, such as feeling doubtful about the feedback or being receptive to criticism (To, 2022). Moreover, other studies assessed the emotional reactions in a longitudinal design and thus laid the focus on the change in students' affect and emotions (Lyu & Lai, 2022; Tay & Lam, 2022). In computer-based learning environments with automated feedback, students often raise concerns about the accuracy and helpfulness of automated feedback (Lipnevich & Smith, 2009b). In their review on feedback perceptions van der Kleij and Lipnevich (2021) emphasize the importance of students being open for feedback and perceiving feedback as useful in order to initiate learning processes. Therefore, in computer-based learning environments, affective feedback engagement is often less about the emotional response to the feedback but rather students' perceived valence of the feedback. Some studies thus assessed perceived characteristics of the feedback such as fairness, helpfulness, or comprehensibility (Seifried et al., 2016), or more generally, positive or negative attitudes towards the feedback (Han, 2017; Han & Hyland, 2015; Z. Zhang & Hyland, 2022).

Behavioral feedback engagement refers to how students implement the feedback they receive. Numerous studies, particularly qualitative case studies, have focused on the operations initiated based on the feedback received (Y. Fan & Xu, 2020; Han & Hyland, 2015; Han & Xu, 2021; Koltovskaia & Mahapatra, 2022; Man et al., 2021; Tian & Zhou, 2020). They argue that, for example, the quality of revision operations (correct, incorrect, or no revision) and observable revision and learning strategies that extend beyond the text itself, such as seeking for additional information or consulting peers and teachers, indicate the level of behavioral feedback engagement. In computer-based learning environments with automated feedback, metrics such as time spent on revising and the number of feedback-revision iterations are also valid measures of engagement, because they indicate the level of student activity in improving their work (Han & Hyland, 2015; Hyland & Zhang, 2018; M. Liu et al., 2017; Z. Zhang & Hyland, 2022). In conclusion, high behavioral feedback engagement is attributed to students who actively improve their texts based on the feedback received, demonstrating a commitment to understanding and incorporating feedback to enhance the quality of their work.

1.3.2 Motivation

Motivation is a critical factor influencing engagement and achievement in learning. Among the theories that explain motivation, expectancy-value theory provides a comprehensive framework, integrating core concepts from various motivational models (Rosenzweig et al., 2019). It offers a strengths-based systems perspective rather than focusing on deficits, making it particularly relevant for understanding the dynamic nature of motivation in educational contexts. According to expectancy-value theory, motivation is shaped by learners' expectancies for success, competence beliefs, the value they assign to a task, and the perceived costs of engaging in it (Wigfield & Eccles, 2000). These factors interact to influence learning behaviors and outcomes. Eccles and Wigfield (2020) conceptually distinguish between time- and task-specific expectancies for success and more stable, domain-specific competence beliefs, acknowledging the theory's historical roots in social-cognitive theory. Expectancies for success involve predicting future outcomes by assessing how well one is likely to perform on an upcoming task. These expectancies reflect one's situation-specific interpretation of their competence regarding the task (Eccles & Wigfield, 2020; Rosenzweig et al., 2022). Competence beliefs, on the other hand, are formed from past performances and represent broad and stable perceptions of one's current abilities (Marsh et al., 2012). In empirical research, competence beliefs are often conducted using either students' domain-specific academic self-concept or self-efficacy (Doménech-Betoret et al., 2017; J. Guo et al., 2016; Jiang et al., 2018; Meyer et al., 2019; Putwain et al., 2019).

Task values refer to the subjective valence of a task, encompassing intrinsic value, utility value, and attainment value. Intrinsic value includes interest, enjoyment, and willingness to participate. Eccles and Wigfield (2020) connect intrinsic value to situational interest (Ryan & Deci, 2020) and intrinsic motivation (Hidi & Renninger, 2006), emphasizing its variability based on the task and time. Utility value refers to the perceived usefulness of a task for achieving one's goals. It is closely linked to extrinsic motivational processes and the incentive-driven aspects of motivation (Rosenzweig et al., 2019; Ryan & Deci, 2020; Schunk et al., 2008). In contrast, attainment value reflects the personal importance and identity-based significance attributed to success in a task (Wigfield et al., 1997). This concept aligns with intrinsic motivational dimensions, where the desire to excel and the personal meaning attached to a task can amplify overall motivation (DeBacker & Nelson, 1999; Elliot & Church, 1997; R. B. Miller et al., 1999; Wigfield & Cambria, 2010b). In contrast to other motivation theories, expectancy-value theory is distinguished by its assumption of a positive relationship between expectancies and values (Rosenzweig et al., 2019). Students who believe they will perform well on a task tend to value it more, while those who doubt their performance may devalue the task to protect their self-worth. Empirical studies consistently support this positive connection between expectancies and values (Durik et al., 2006; Meyer et al., 2019; Wigfield et al., 1997; Wigfield & Eccles, 2002).

Cost refers to the perceived negative aspects of engaging in a task, such as emotional stress, anxiety, or fear of missing out. These factors shape decision-making and often act as barriers to engagement. Eccles and Wigfield (2020) emphasize that every learning activity entails costs, which interact with perceived benefits to form a cost-benefit ratio. For example, costs influence behavior through prioritizing competing tasks or goals, as individuals direct their attention to activities they view as more important, urgent, or rewarding, often at the expense of other tasks. Additionally, the reluctance to invest effort arises when individuals perceive the effort required as disproportionate to the potential benefits, prompting them to disengage or avoid the task altogether (Inzlicht et al., 2018; Kool et al., 2010). Finally, fear of failure can further reduce engagement, as individuals may avoid tasks to protect their emotional well-being or self-esteem when they perceive a high risk of underperforming (Conroy, 2004; J. Lee et al., 2013). These motivational dynamics evolve over time and are shaped by the immediate context of each decision. Eccles and Wigfield (2020) highlight that expectancies for success, task values, and costs are highly situation-dependent. As a result, both conscious and subconscious choices are influenced by current circumstances, reflecting the fluid interaction of these factors in the decision-making process.

Each construct of expectancy-value theory particularly predicts certain academic outcomes. Competence-related beliefs are strongly associated with aspects of performance and achievement (Baadte & Schnotz, 2014; Bong et al., 2012; Meece et al., 1990; Safavian, 2019; Safavian & Conley, 2016; Taboada et al., 2009). In contrast, task values more strongly predict aspects of task engagement and choices, such as self-regulation, feedback engagement, and feedback seeking (Gan, Mark J. S. & Hattie, 2014; Iraj et al., 2021; Yossatorn et al., 2024), and also more long-term decisions like selecting college majors, courses, or extracurricular activities (Eccles & Wigfield, 2020; Rosenzweig et al., 2019). Focusing on specific components of task value, intrinsic value is a strong predictor of daily academic engagement. Students who perceive higher intrinsic value in a subject are more likely to engage in related activities, put effort into courses, and attend class (Durik et al., 2006; J. Guo et al., 2016; Watt et al., 2012). Dietrich et al. (2017) showed a reciprocal relationship between effort and intrinsic value. Students invested more effort in situations where they valued the task highly but also had higher intrinsic value when they had invested more effort in previous situations. Utility value contributes to the development of interest in a specific subject or task and fosters deeper engagement and persistence with a task and related feedback (Hulleman & Harackiewicz, 2020). For example, studies by Hulleman et al. (2010) and Canning and Harackiewicz (2019) found that utility value interventions positively impacted interest and performance, particularly for students who expected or had previously shown low performance. Similarly, Hecht et al. (2021) reported overall positive effects of utility value interventions on interest development, especially for less confident students and those who had already demonstrated an initial interest in a topic. Attainment value is particularly influential in forecasting long-term academic participation. Research by Durik et al. (2006), Updegraff et al. (1996), and Watt et al. (2012) found that a composite of attainment and utility value was more predictive of high school students' career intentions than intrinsic value. Moreover, Safavian et al. (2019; 2016) reported that perceived attainment value influenced some students' course-taking decisions, whereas perceived utility value did not. There has been less research on perceived cost, but existing studies indicate that it predicts performance and avoidance-related behaviors or emotions, such as academic cheating, procrastination or dropout intentions (Barkela et al., 2024; Flake et al., 2015; Jiang et al., 2018; Perez et al., 2014).

In computer-based learning environments, considering students' perception of their competence beliefs, values, and costs can be elucidating as they determine students' engagement and persistence with digital content and automated formative feedback (Jung & Lee, 2018; Seifried et al., 2016). When students believe they can succeed and find the learning tasks meaningful, they are more likely to engage deeply and achieve better outcomes (Wigfield & Cambria, 2010a). Moreover, automated formative feedback can manipulate both these components by providing immediate assessments of performance, elucidating potential improvement, and highlighting the relevance of tasks like summarizing. Students may adapt their expectancies and competence beliefs when receiving formative feedback because they may identify clear indicators of progress and areas for improvement. Additionally, having included several value components, expectancy-value theory allows for an in-depth analysis of the interplay between those components and their impact on students' feedback engagement and performance (Rosenzweig et al., 2019). For example, in the context of this dissertation, being informed about the value students place on summarizing and their ability beliefs of summarizing might inform about their engagement with the texts and the automated formative feedback provided in a computer-based learning environment.

Eccles and Wigfield's (2020) expectancy-value theory aims to provide a comprehensive, integrative framework that unites insights from various motivation theories while focusing on strengths rather than deficits. This approach resulted in a multi-faceted model that has become foundational for understanding educational motivation across diverse contexts. Numerous studies have explored motivational prerequisites for learning (J. Guo et al., 2016; J. Lee et al., 2013; Meyer et al., 2019; Nagengast et al., 2011; Trautwein et al., 2012), the potential to influence expectancies and values through interventions (Acee et al., 2018; Gaspard et al., 2015; Hecht et al., 2021; M. L. Johnson & Sinatra, 2013; Linnenbrink-Garcia et al., 2018; Perez et al., 2022; Rosenzweig et al., 2020), their development over time (Dietrich et al., 2017, 2019; Han, 2017; Marsh et al., 2016; Perez et al., 2019), and their effects on learning processes, achievements, and engagement (Alipio, 2020; Durik et al., 2006; W. Fan &

Williams, 2010; Jiang et al., 2018; Putwain et al., 2019; Sun & Rueda, 2012). Additionally, research has examined the interaction between expectancies, values, and cognitive processes, such as the investment of (mental) effort (Dietrich et al., 2017; Marsh et al., 2016).

1.3.3 Cognitive Load

Engaging successfully in complex tasks like summarizing poses a certain amount of cognitive load on the learners as they construct and assimilate new knowledge with their cognitive schemas (Sweller, 2020). In computer-based learning environments, the management of cognitive load is crucial for optimizing learning outcomes. Cognitive load theory posits that learners have limited capacity for processing information, which can be divided into intrinsic, extraneous, and germane cognitive loads (Sweller, 2011). Intrinsic load is related to the complexity of the material itself, extraneous load pertains to how the information is presented, and germane load refers to the cognitive resources dedicated to process, construct, and automate schemas (Kalyuga & Singh, 2016; Paas et al., 2003; Schmeck et al., 2015). Effective computer-based learning environments are designed to reduce extraneous load and optimize germane load to enhance learning efficiency (Plass et al., 2010). Tools such as adaptive automated feedback systems, which provide support based on the learner's performance, can help manage cognitive load by ensuring that tasks remain within the learner's zone of proximal development (Vygotsky, 1978). Additionally, providing multiple representations of information can enhance learning by supporting dual coding and thus reducing cognitive load (Mayer, 2014).

An established approach to assess cognitive load is to ask learners about their invested mental effort (Paas et al., 1994, 2003; Schmeck et al., 2015; Sun et al., 2019). Studies have shown that in order to achieve a learning goal, students anticipate the effort they will have to invest to master the mental load induced by a learning task, while considering the perceived task difficulty (Feldon et al., 2019; Salomon, 1983). However, humans inherently perceive mental effort as costly, prompting a tendency to minimize its expenditure (Kool et al., 2017; Shenhav et al., 2017; Yee & Braver, 2018). Kool et al. (2010) demonstrated that individuals consistently opt for actions associated with lower cognitive demands. This behavior is corroborated by Gieseler et al. (2020), who observed that individuals tend to select less effortful tasks after initially exerting significant mental effort. Nonetheless, the potential for rewards, the perceived value of the task, and the anticipated success of task performance can enhance the willingness to invest mental effort (Dietrich et al., 2017; Frömer et al., 2021; Manohar et al., 2015; Marsh et al., 2016). Contrary, if the task does not seem aligned with students' personal goals or interests, their motivation to invest mental effort may decline (Paas et al., 2005; Yee & Braver, 2018). Moreover, if a task is seen as too simple or excessively difficult, learners may also become disengaged and less inclined to expend the necessary cognitive resources (Paas et al., 2005).

Automated feedback systems can play a significant role in managing mental effort. Interactive features such as personalized formative feedback and prompting students to reflect on the text and their learning processes can engage learners in active learning, thereby promoting deeper cognitive processing and better retention of information (Chi, 2009). Moreover, immediate and specific feedback helps learners correct errors and understand concepts more deeply, which can lead to more efficient schema construction (Hattie & Timperley, 2007). Additionally, *formative* feedback and prompts directed at the cognitive processes of summarizing may encourage students to invest more mental effort and engage more deeply with the task and the related feedback (Graham et al., 2015; B. W. Miller, 2015; Nückles et al., 2020; Sun et al., 2019).

Thus, mental effort and feedback engagement may reciprocally affect each other. Additionally, reciprocal relationships have also been shown between mental effort and motivation (Capa & Audiffren, 2009; de Araujo Guerra Grangeia et al., 2016; Dietrich et al., 2017; Marsh et al., 2016; Paas et al., 2005; Yee & Braver, 2018). Furthermore, motivation highly predicts engagement in formative feedback (Durik et al., 2006; J. Guo et al., 2016; Watt et al., 2012). However, systematic investigations of the reciprocal relationships between mental effort, motivation, and feedback engagement in automated feedback systems are scarce (Han, 2017).

1.4 Instructional Factors

For designing and evaluating automated formative feedback, instructional factors play a crucial role. Among these, the type of task significantly shapes how feedback is perceived and utilized by learners. Summarizing, as a generative task, has unique implications as it requires students to process information deeply, identify key ideas, and reorganize content in their own words, making it both cognitively demanding and highly beneficial for developing academic skills (Dunlosky et al., 2013; Fiorella & Mayer, 2016; E. Kintsch, 1990). To support such tasks effectively, another important instructional factor is the use of prompts (Ahn, 2022; Green & Holman, 2021; Ko, 2009; Wischgoll, 2017). Prompts can scaffold the learning process by guiding students through specific strategies, potentially complementing the automated feedback and fostering more structured engagement with the task.

Prompts are frequently used communication tools that bridge the gap between learners and digital platforms in computer-based learning environments. Designed to guide and support student interaction with content, prompts enhance learning by targeting cognitive, metacognitive, and motivational processes. They are widely used to support various aspects of learning, including knowledge acquisition (Hattie et al., 1996; L. Zheng, 2016), writing (Proske et al., 2012; Zellermayer et al., 1991), and summarizing (Ahn, 2022; Green & Holman, 2021). Furthermore, prompts can encourage reflection on the learning process and outcomes (Bannert, 2006; Krause & Stark, 2010), promote self-regulation and metacognition (Engelmann & Bannert, 2021; Lim et al., 2023; Teng, 2022), and guide strategy use during task processing (Lehmann et al., 2019; Proske et al., 2012; Roscoe & Chi, 2008). A wide range of studies has investigated the effectiveness of such prompts (Chen & Huang, 2014; Delen et al., 2014; Engelmann et al., 2021; Hefter et al., 2023; A. M. Johnson et al., 2011; Krause & Stark, 2010; Lehmann et al., 2019; Roelle et al., 2015; Teng, 2022; Wischgoll, 2016; Yang et al., 2021) and associations between prompts and individual dispositions like motivation (Gidalevich & Kramarski, 2019; Lehmann et al., 2014; Nückles et al., 2010; Wäschle et al., 2015) and self-efficacy (Gentner & Seufert, 2020; Moos & Azevedo, 2008; Müller & Seufert, 2018).

Prompts in summarizing tasks can be designed to activate key cognitive processes and support self-regulated activities (Nückles et al., 2020). *Activating prior knowledge* is crucial for anchoring new information to existing cognitive schemas, making the integration of new content more efficient (Gurlitt & Renkl, 2010; Mayer, 1997; Piaget, 1950). For example, a prompt might present the heading of a text and ask students what they already know about this topic. Furthermore, prompts can encourage learners to *identify relevant information* for inclusion in a summary and filter out irrelevant details, enhancing their ability to distill complex information into a coherent summary (Kellogg & Raulerson, 2007; Westby et al., 2010). This process requires critical thinking and discernment which can be stimulated by prompts that ask learners to formulate questions answered in the text or to create headings for key aspects. Moreover, *aligning new information* with prior knowledge ensures that learners effectively *integrate new concepts* into existing cognitive schemas, solidifying their understanding by embedding it within a broader knowledge context (Becker-Mrotzek et al., 2014; Hidi & Anderson, 1986). This integration can be encouraged by prompts that ask students to find examples illustrating key concepts or to clarify unclear, challenging, or unfamiliar passages.

Additionally, prompts can guide learners in *constructing a mental model* of the text, helping them organize and understand the content more coherently, which is crucial for both comprehension and recall (E. Kintsch, 1990; Leopold et al., 2019). Questions that ask about the text's line of thought or overall structure can facilitate this process. Finally, the step of *translating this mental model into written text, using their own words*, requires learners to clearly and concisely articulate their understanding (E. Kintsch et al., 2000; Perin et al., 2017). This final step in summarizing reinforces deep cognitive engagement and ensures that learners have thoroughly processed the information, ultimately enhancing their comprehension and retention.

For support in self-regulated activities, prompts can support the *planning* phase by encouraging students to set goals, select strategies, and organize resources before engaging with the task (Kellogg, 1988; Kellogg & Whiteford, 2012). In the context of summarizing, effective planning helps learners to structure their approach, focus on key concepts and organize their thoughts effectively, leading to a more coherent and accurate summary (A. L. Brown et al., 1983). The planning phase can be encouraged through prompts that ask learners to identify the main points that need to be covered in their summary. Furthermore, prompting to *monitor* one's learning process involves the continuous assessment of one's understanding and progress while working on a task (de Silva & Graham, 2015; Zimmerman & Kitsantas, 2007). During summarizing, monitoring enables learners to ensure they are staying on track, following their plan, and accurately capturing the key points of the material. This ongoing process helps in identifying any missed details or misunderstandings, allowing for necessary adjustments. Prompts designed to support monitoring aim to cultivate habits of self-assessment and reflection. For instance, learners might be prompted to re-read certain passages to clarify aspects that remain unclear. Finally, prompts may address *evaluating* one's product, where learners assess the quality and effectiveness of their work upon task completion

(Panadero et al., 2017). In summarizing, evaluating involves reviewing the summary to ensure that it accurately represents the key ideas, is well-organized, and adheres to the predefined criteria (in this dissertation: content coverage, avoidance of copied words, redundancy avoidance, relevance, length). Prompts for evaluating encourage a reflective mindset, urging learners to critically assess their performance and learn from the experience (Engelmann et al., 2021). For example, learners might be asked whether all main points have been included and whether redundancy has been effectively avoided.

Research has shown that prompting a combination of both cognitive and self-regulatory strategies enhances learning outcomes more than prompting either one of them alone. For example, Berthold et al. (2007) found that learners who received either cognitive prompts or a combination of cognitive and self-regulatory prompts learned more than those without prompts, while self-regulatory prompts alone did not lead to increased learning and resulted in the use of fewer cognitive strategies. In a replication study, Nückles et al. (2009) showed that self-regulatory prompts alone could improve learning if accompanied by opportunities for planning and applying remedial strategies to improve comprehension. However, the most effective results still came from combining cognitive and self-regulatory prompts. Similarly, Glogger et al. (2012) found that learners who used both cognitive and self-regulatory strategies outperformed those who used only one type, reinforcing the significance of prompting a balance of these strategies for optimizing learning outcomes. In addition, Roelle et al. (2017) found that the sequencing of cognitive and self-regulatory prompts also affects learning outcomes, with self-regulatory prompts provided first leading to better organization and comprehension of content. Learners who engage in self-regulatory processes, such as identifying gaps and planning remedial strategies before cognitive processing, may develop a stronger knowledge base, allowing for deeper content elaboration and organization.

Moreover, the effectiveness of prompts differs depending on students' levels of strategy sophistication and the development of their educational identities, e.g., how they perceive themselves as learners and interact with support such as prompts. For students with lower competencies, prompts can model effective summarizing strategies and activate cognitive processes, compensating for gaps in their abilities and guiding them through processes they might not yet have mastered (Gentner & Seufert, 2020). These students often benefit from structured guidance, as prompts can help scaffold their understanding and provide a framework for organizing their thoughts. Whereas learners who have developed robust self-regulatory strategies may use prompts as a reinforcement of their existing approaches (Engelmann et al., 2021). However, for high-performing students, prompts may sometimes be perceived as disruptive or redundant, potentially interfering with their more autonomous and sophisticated learning processes (cf. Gidalevich & Kramarski, 2019; Kalyuga, 2014; Nückles et al., 2010).

Few studies are concerned with the comparison of the effects of automated formative feedback, instructional prompts, or a combination of both on learning outcomes. In two studies by Krause and Stark (2010) both reflection prompts and immediate elaborated feedback enhanced learning outcome. Yet, only reflection prompts influenced students' perception of having reflected on their learning. Similarly, van den Boom (2004, 2007) found that reflection prompts stimulated students to engage in reflective activities and that additional tutor feedback enhanced these reflections. However, in their studies, prompts and feedback were provided by peers or human tutors, with the primary learning goal focused on acquiring self-regulation strategies. The interplay of instructional prompts and automated formative feedback without the moderation of humans in a computer-based learning environment about summarizing has not yet been explored. However, for pedagogical considerations in designing effective learning environments and for deeper understanding learning processes in automated feedback systems, it is necessary to examine how instructional prompts and automated formative feedback influence learner behavior and skill acquisition.

2. The Present Dissertation

The previous sections have highlighted that automated formative feedback systems allow teachers to effectively provide feedback instantly, simultaneously, and frequently without creating an excessive workload. Hence, with the support of technology, new opportunities arise to foster university students’ summarizing skills while easing some of the teachers’ time demands associated with providing formative feedback, particularly in large courses. However, effectively integrating automated formative feedback systems into university teaching practices requires a comprehensive understanding of technology-enhanced feedback processes. While there already is research on this topic, critical gaps were identified that warrant further investigation of feedback characteristics (1), individual learner factors (2), and instructional factors (3).

1) With regard to the characteristics of feedback, several studies showed that summarizing can be supported with automated formative feedback which informs about certain aspects such as content coverage, avoidance of copied words, redundancy avoidance and adhering to short length (e.g. Lenhard et al., 2012; Wade-Stein & Kintsch, 2004). However, most studies employing such a feedback system have been conducted in elementary and middle school settings (Lenhard et al., 2012; Sung et al., 2016; Wade-Stein & Kintsch, 2004) aiming at fostering reading comprehension. Introduced to the university context, it is unclear whether such an automated formative feedback system can promote summarizing skills in university students, as well. Furthermore, it is unclear how this feedback system supports the fulfilment

of the various task criteria (full content coverage, avoidance of copied words, redundancy avoidance, relevance, and length). Following, a first research question is:

> *Research question 1.1: How effective is automated formative feedback in promoting the development of summarizing skills in university students over time?*

One advantage of automated feedback systems is, furthermore, the formative provision of real-time feedback in several loops. For example, Sung et al. (2016) operationalized the number of feedback loops as tool use to determine whether students have been relying on the tool or enhanced their skills. However, a systematic examination of the effects of *formative* feedback and performance such as summary quality has not yet been investigated. Thus, another research question is concerned with:

> *Research question 1.2: How does the formative nature of the automated feedback affect performance improvement?*

2) Regarding individual learner factors, feedback engagement is a crucial factor, as the quality of automated feedback processes largely depends on how recipients utilize the feedback to its full potential. Currently, research on feedback engagement has primarily happened in the context of EFL (Y. Liu & Sun, 2021; M. Liu et al., 2017; Hyland & Zhang, 2018; also cf. Pearson, 2024) and focused on qualitative investigations through case studies (Han & Hyland, 2015; Han & Xu, 2021; Z. Zhang, 2020). Longitudinal research and studies that thoroughly examine the different dimensions of feedback engagement are scarce (Wang, 2014; Yu, Zhang, et al., 2019). Thus, several questions remain unanswered: How do students engage with the automated formative feedback over several sessions? Does previous feedback engagement affect subsequent engagement? How are the dimensions of feedback

engagement connected? To examine learners' engagement with feedback in an automated formative feedback system for promoting summarizing, the following research question was formulated:

Research question 2.1: How consistent are students' cognitive, affective, and behavioral feedback engagement patterns across several time points?

Furthermore, research on individual factors influencing feedback engagement has largely focused on specific aspects, resulting in a limited understanding of the broader dynamics. Studies have examined the impact of students' perceptions (Ali et al., 2018; Carless & Boud, 2018) and motivation (Malecka & Boud, 2021) on feedback engagement. Similarly, effort has been explored in the context of general learning engagement (Putwain et al., 2019) and in relation to reciprocal motivational influences (Dietrich et al., 2017; Marsh et al., 2016). However, the specific role of mental effort in automated formative feedback engagement remains uninvestigated. Moreover, the reciprocal relationships between mental effort, motivational resources, and feedback engagement, particularly over time, are still underexplored. These gaps raise critical questions: How do learner factors like mental effort and motivation shape feedback engagement? Are there reciprocal relationships between these factors and feedback engagement over time? Thus, the following research question is:

Research question 2.2: How do the reciprocal relationships of cognitive and motivational resources impact feedback engagement processes in the perspective of time?

3) Moreover, the explicit prompting of summarizing strategies has been shown to be an important instructional factor (Roscoe & Chi, 2008; Wischgoll, 2017). For example,

explicitly prompting summarizing strategies can be a complementary approach, as students may become more aware of these strategies, adopt them, and be able to intentionally use them when needed (Ahn, 2022; Ko, 2009). However, the effects of offering prompts compared to or in combination with automated formative feedback has not yet been systematically investigated. Thus, several questions arise: How do instructional prompts targeting summarizing strategies effectively support summarizing compared to automated formative feedback? Can such prompts offer an additional benefit compared to feedback alone? Do these prompts and automated formative feedback activate different activities in the summarizing process? Hence, the following research questions were formulated:

> *Research question 3.1: How can instructional prompts support the acquisition of summarizing skills compared to automated formative feedback?*

> *Research question 3.2: How does a combination of instructional prompts and automated formative feedback affect the acquisition of summarizing skills?*

To address these research questions, a computer-based learning environment with automated formative feedback to foster summarizing skills was developed and evaluated in three studies based on the three perspectives for designing feedback strategies (cf. Narciss, 2013). The first study examined the effectiveness of automated formative feedback in improving students' ability to summarize. The second study considered learners' individual factors, the reciprocal relationships between mental effort, motivation, and feedback engagement. The third study investigated the interplay of instructional prompts and automated formative feedback in enhancing summarizing skills.

All three studies were longitudinal, conducted at six measurement points, with undergraduate elementary education students. For each measurement point students were asked to summarize a pedagogical scientific text about student-teacher interaction. The texts were selected to have similar levels of text difficulty, as determined by a readability index for German texts (LIX; Lenhard & Lenhard, 2014). At the first and last measurement points, participants submitted their summary but did not receive feedback. For the other four sessions, they could write their summary, upload it, and receive automated formative feedback up to ten times, including semantic feedback and score feedback. Semantic feedback marks copied, irrelevant, and unknown words and redundant sentences in the students' summary and details the redundant sentences in a pop-up window, listing similar sentences by color. In percentages as horizontal bars, score feedback shows the extent to which students' summaries meet the task criteria of full content coverage, avoidance of copied words, avoidance of redundancy and length not to exceed 30% of the original text. A detailed description of the learning environment can be found in manuscript 1.

For all studies, students were asked about their language proficiency, and socio-demographic data at the beginning of the first task. Time on task, and the number of feedback loops were monitored during each task processing. For each summary, single aspects of and overall text quality was assessed with LSA (Landauer, 1999). For the first study, an additional control group with special education students was collected who summarized the same texts as the experimental group but did not receive any feedback or comments on their summaries. For the second study, additionally students' motivational and cognitive resources as well as students' feedback acceptance were assessed at different time points: Students' self-concept, and utility/attainment value of summarizing was assessed at the beginning of the first task, students' expectancy for success, intrinsic value, and cost was assessed before each task, and students' mental effort, and feedback acceptance was assessed after each task. For the third study, additional experimental groups were collected with one group receiving prompts for applying summarizing strategies instead of automated formative feedback and one group receiving these prompts in addition to the automated formative feedback.

The results of these studies have been reported in three manuscripts (Manuscript 1: Barkela & Leuchter, 2024a; Manuscript 2: Barkela et al., 2023; Manuscript 3: Barkela & Leuchter, 2024b) that have been submitted to or published in peer-reviewed journals and are in the appendix of this dissertation. The methods and key results of the studies are summarized in the following chapters (chapter 2.1–2.3), a detailed description of the studies can be found in the respective manuscripts. For an overview of the aims, methods and findings of the articles, see Table 3.

Table 3
Overview of the Three Presented Manuscripts

	Manuscript 1 **Effectiveness of Automated Formative Feedback**	**Manuscript 2** **The Impact of Cognitive and Motivational Resources on Engagement with Automated Formative Feedback**	**Manuscript 3** **The Interplay of Instructional Prompts and Automated Formative Feedback**
Topic	Effectiveness of automated formative feedback in improving students' ability to summarize	Reciprocal relationships between cognitive and motivational resources and automated formative feedback Engagement	Comparison of instructional prompts, automated formative feedback, or a combination of both as support for acquiring summarizing skills
Aims	1.1) Investigating the effectiveness of automated formative feedback in improving summarizing skills. 1.2) Examining each aspect of summarizing (content, length, avoidance of copied words, redundancy avoidance) and their specific promotion by the automated formative feedback. 1.3) Exploring the impact of feedback loops on the improvement of summarizing skills.	2.1) Examining the development (indicated by time) of feedback engagement. 2.2) Investigating the reciprocal relationships of feedback engagement with cognitive and motivational resources	3.1) Investigating the effectiveness of instructional prompts versus automated formative feedback in improving summarizing skills. 3.2) Exploring if combining both prompts and feedback offers additional benefits in enhancing summarizing skills.
Experimental variables	– Time – Number of feedback loops	– Expectancy for success, self-concept – Utility/ attainment, intrinsic value – Cost – Mental effort (T; T-1) – Performance	– Three different forms of support: prompts, feedback, and a combination of both – Time, time on task – Language proficiency – Text difficulty
Outcome variables	– Text quality – Content – Length – Avoidance of copied words – Redundancy avoidance	– Cognitive feedback engagement – Affective feedback engagement – Behavioral feedback engagement	– Text quality – Number of feedback loops
Sample	138 B. Ed. Elementary Education students / M. A. Special Education students	330 B.Ed. Elementary Education students	254 B.Ed. Elementary Education students

2.1 Effectiveness of Automated Formative Feedback

The first manuscript (Barkela & Leuchter, 2024a) was concerned with the development of an expert computer-based learning environment (*FALB*) with automated formative feedback for promoting summarizing. First, design principles and criteria for summarizing were derived, emphasizing task features that promote the processing of texts. Key criteria for effective summaries included content coverage, word copying avoidance, redundancy avoidance, and reduced length (Wade-Stein & Kintsch, 2004). Additionally, task designs such as preventing simultaneous access to the text while writing and setting length limits were highlighted for their role in fostering concise and accurate summaries (Hidi & Anderson, 1986; Hill, 1991). Moreover, formative feedback was identified as relevant for facilitating iterative revisions, promoting deep text processing, and ensuring adherence to task criteria aligned with an external standard (Graham, 2018; Kellogg & Raulerson, 2007). Second the computer-based learning environment *FALB* was introduced, detailing its structure and the automated feedback system based on LSA. Moreover, a composite score was developed and validated to holistically assess the text quality of summaries.

Regarding research question 1.1, the study examined whether automated formative feedback effectively improved summarizing skills by analyzing the experimental group's text quality across six measurement points compared to a control group. In addition to overall text quality, changes in single aspects of text quality were explored to identify which aspects of summarizing were most influenced by automated feedback. Lastly, addressing research question 1.2, the study investigated how the number of feedback loops affected summary text quality, drawing on prior findings that emphasize the importance of iterative revisions and feedback in the learning process (Bereiter & Scardamalia, 1987; J. A. Butler & Britt, 2010).

To address these research questions, a study was conducted with 138 participants (N = 87 B.Ed. elementary education students in the experimental group, N = 51 M.A. special education students in the control group). The experimental group used FALB to summarize the six scientific texts, while the control group summarized the same texts in a computer-based environment without automated feedback. Results showed that interacting with FALB significantly fostered summarizing skills. Automated formative feedback led to improvements in text quality, aligning internal and external reference standards (cf. Narciss, 2017). Additionally, the automated formative feedback may have fostered consistent monitoring and engagement, sustaining high performance even in the final session without feedback, that suggests an effective transfer of feedback insights.

The feedback particularly supported adherence to predefined length and the avoidance of word copying while maintaining high content coverage. These cognitive processes are essential for constructing a mental model of the text. However, redundancy avoidance was not improved, possibly reflecting unelaborated prior knowledge and insufficient skills in writing concisely. Limited prior knowledge may hinder the ability to infer, condense the gist, and reorganize ideas, making concise summaries challenging to produce. Finally, a positive relationship was observed between the number of feedback loops and text quality, indicating that frequent revisions enhanced summary quality.

Two questions emerge from the discussion of the results in the first manuscript. First, while engaging in more feedback loops leads to better outcomes, there are interindividual differences in how effectively learners utilize formative feedback. This underscores the need to investigate feedback engagement in relation to individual prerequisites. Given the communicative and interactive nature of feedback processes, particularly in automated systems where the learners' willingness to engage with feedback is the driving force, it is essential to explore both feedback engagement and the individual factors influencing it. This is addressed in the second study of this dissertation which investigates relationships between engagement with automated formative feedback and learners' prerequisites (see chapter 2.2).

Second, the observed low levels of redundancy avoidance suggest that while automated formative feedback effectively supported the revision phase of writing, it may have overlooked the planning phase. As a result, students might not have developed strategies to condense core aspects of the text and restructure ideas—key activities typically emphasized during planning. This gap is considered in the third study of this dissertation that incorporates support for the planning phase and compares different forms of support (see chapter 2.3).

2.2 Impact of Cognitive and Motivational Resources on Feedback Engagement

The second manuscript (Barkela et al., 2023) examined the consistency of and reciprocal relationships between cognitive, affective, and behavioral feedback engagement, and cognitive and motivational resources. Research has shown that automated formative feedback can promote summarizing skills effectively (Chew et al., 2019; Lenhard & Lenhard, 2014; Sung et al., 2016). However, its effectiveness highly depends on the individual engagement with feedback (Handley et al., 2011; Price et al., 2011). Potential individual factors that may affect the level of feedback engagement are cognitive and motivational resources (Han, 2017). Addressing research questions 2.1 and 2.2 of this dissertation, suitable measures were theoretically derived for 1) cognitive, 2) affective, and 3) behavioral feedback engagement and specified a model with direct, indirect, and lagged effects which integrated the three dimensions of feedback engagement and related them to cognitive and motivational resources in the perspective of time.

To test the model, a study was conducted with 330 B.Ed. elementary education students. Over a four-session intervention focused on summarizing texts, students' motivational resources were assessed using the expectancy-value theory (Eccles & Wigfield, 2020), while cognitive resources were evaluated based on the level of invested mental effort (Naismith et al., 2015; Paas, 1992). Previous performance was measured as the final text quality score from the previous session. Absolute model comparisons indicated a good fit.

First, the interaction among the three dimensions of feedback engagement was examined. Cognitive feedback engagement was positively related to affective and behavioral feedback engagement, while affective feedback engagement showed a negative association with behavioral feedback engagement. Moreover, previous affective and behavioral feedback engagement strongly predicted subsequent engagement in these dimensions. However, no significant correlation was found between previous and subsequent cognitive feedback engagement. Second, the relationships between the three dimensions of feedback engagement and previous performance, as well as cognitive and motivational resources, were analyzed. 1) Cognitive feedback engagement was positively associated with previous performance but showed no significant associations with cognitive or motivational resources. 2) Affective feedback engagement was positively associated with intrinsic value but negatively related to situational expectancies, invested mental effort, and previous performance. 3) Behavioral feedback engagement exhibited a positive association with situational expectancies and invested mental effort. These findings suggest the importance of encouraging students to allocate greater mental resources to seeking feedback more frequently, thereby enhancing their behavioral engagement in the learning process.

2.3 Interplay of Instructional Prompts and Automated Formative Feedback

The third manuscript (Barkela & Leuchter, 2024b) focused on the effects of different forms of support and related pedagogical considerations. Research has shown that instructional prompts are a suitable support in computer-based learning environments as they can activate the use of effective summarizing strategies (Ahn, 2022; Green & Holman, 2021; Leopold et al., 2019). Likewise, research has emphasized the important role of individualized formative feedback (Graham, 2018; Schunk & Rice, 1991). Yet, a systematic comparison of supporting summarizing skills through instructional prompts versus automated formative feedback, as well as the complementary effects of combining these approaches, has not yet been conducted (cf. Krause & Stark, 2010; van den Boom et al., 2004, 2007). Therefore, addressing research question 3.1 of this dissertation, the study investigated the effectiveness of instructional prompts compared to automated formative feedback in improving summarizing skills. Additionally, it examined whether combining prompts and feedback offers further benefits in enhancing these skills and analyzed their complementary effects, as addressed in research question 3.2.

For this purpose, a study was conducted with 254 B.Ed. elementary education students (feedback group: $N = 87$, prompt group: $N = 75$, combi group: N = 92). The feedback and combi groups summarized six scientific texts using the computer-based learning environment *FALB*, while the prompt group summarized the same texts in a learning environment where instructional prompts were provided but without external feedback. To ensure the validity of the findings, external influences such as time on task, language proficiency, and text difficulty were controlled for. Results showed that instructional prompts alone supported the development of summarizing skills, albeit to a lesser extent than automated formative feedback. Combining the two did not yield significant additional benefits. This finding may suggest a ceiling effect, indicating that learning potential is maximized with automated formative feedback. However, students in the combi group requested slightly less formative feedback, suggesting a possible substitution effect, where prompts reduced their reliance on feedback. Interestingly, and contrary to prior studies (cf. Manwaring et al., 2017; Z. Zhang & Hyland, 2022), neither time on task nor text difficulty influenced the number of feedback loops, pointing to other factors as key drivers of feedback engagement.

3. General Discussion

The aim of this dissertation was to explore whether and under what conditions automated formative feedback can effectively enhance summarizing skills among university students. This is important because, promoting summarizing at university can support students' academic development, reading comprehension, and writing competency (Dunlosky et al., 2013; J. Li, 2014; Mok & Chan, 2016; Schoonen, 2019). However, academic staff often has not enough resources to offer immediate, simultaneous, and frequent feedback for these fundamental skills. Computer-based learning environments with automated formative feedback can help alleviate teachers' workload while supporting students to develop their skills in summarizing. However, as discussed earlier in this dissertation, several research gaps remained so far leading to important research questions (cf. pp 8f. & 35ff.). To address these, automated formative feedback processes were investigated in three studies, focusing on the following perspectives: characteristics of the feedback, individual learner factors, and instructional factors (Narciss, 2013, 2017). In the following, the three studies and their results will be discussed regarding their potential to answer the research questions and their overall contribution to the field. Moreover, a critical discussion of the limitations and an outlook on future research perspectives will be given. For an overview of the results presented in the articles, see Table 4 (p. 66).

3.1 Discussion of the Research Questions

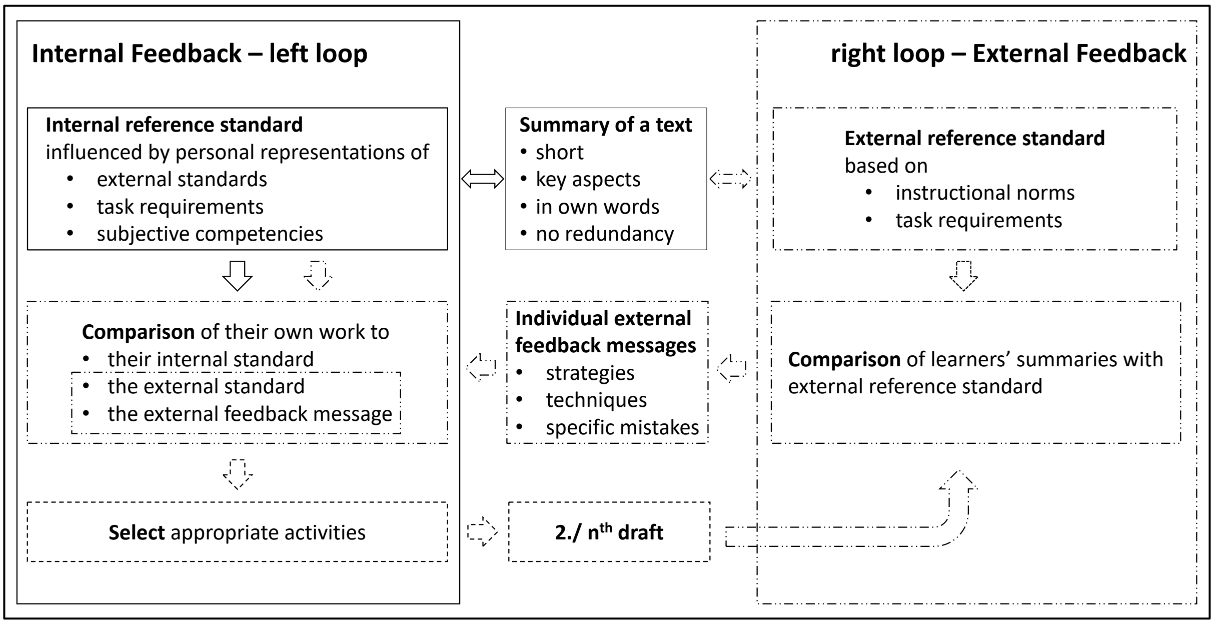
3.1.1 Research Question 1.1: How effective is automated formative feedback in promoting the development of summarizing skills in university students over time?

To examine the effects of automated formative feedback, a feedback algorithm based on LSA was implemented in an online learning environment for summarizing in an elementary education bachelor's program. The first study of this dissertation was concerned with the effectiveness of automated formative feedback provided over several sessions compared to a control group. Students who received automated formative feedback were more likely to improve their summaries' text quality over the six sessions than students in the control group who had no support and did not improve. Accordingly, in the third study, participants in the experimental group who received automated formative feedback again significantly improved their summaries' text quality over the six sessions. These results corresponds to previous findings which attribute a high degree of effectiveness to automated formative feedback (Lenhard et al., 2012; Wade-Stein & Kintsch, 2004; cf. Lenhard, 2008).

In the first study, the control group wrote their summaries based solely on their internal understanding of what constitutes an effective summary. According to the ITF model (Figure 2), since the students did not receive any external support or feedback, they relied solely on internal feedback, generated within the model's left loop. This loop represents the self-regulatory cycle, where learners engage in internal information processing and internal assessment like monitoring their understanding of the text and evaluating their summaries against personal standards, without external input. Although the task criteria were communicated, the students' interpretation of these criteria was shaped by their individual reference standards. As a result, their ability to assess the gap between their current performance and the desired outcome was limited by these personal benchmarks, lacking an external reference to guide their evaluation. Consequently, without external feedback to trigger adjustments in their approach, the students' potential to acquire more advanced summarizing skills or improve their performance was constrained. Any improvement in their summaries would likely occur at random and only if they happened to reinterpret the task criteria and adjust their internal reference standards in subsequent sessions. However, such a reinterpretation without external impulses is rather unlikely (Vosniadou, 2008).

Figure 2

Simplified ITF Model Applied to Summarizing Tasks (Narciss, 2006, 2013)



In the experimental group, students received external feedback, which activated both loops of the ITF model (Figure 2). Unlike the control group, who could rely solely on internal feedback in the left loop, the experimental group benefited from external feedback, allowing interaction between both the left and right loops of the model. The external feedback consisted of semantic feedback, which explicitly marked phrases needing improvement. These cues activated the left loop and encouraged students to repeatedly evaluate their drafts against the external reference standard and adjust their internal standards accordingly. Furthermore, students engaged in internal cognitive processes such as monitoring their understanding of the text and self-evaluating their summaries. However, unlike the control group, their internal reflections were continuously informed by external feedback, enhancing their ability to self-regulate. In the right loop, the external feedback provided a reference point, set by the expert summary, against which students could measure their performance. Along with semantic feedback, students received score feedback that illustrated their performance levels in horizontal bars in percentages, making the external standard visually clear. This visualization allowed students to directly compare their internal representations of the task criteria and their drafts with the external benchmark. As a result, they could more easily identify gaps between their actual performance and the desired outcomes, enabling them to revise their summaries accordingly.

The first manuscript investigated how effectively students improved across individual criteria, aiming to understand how automated feedback promotes summarizing skills. The results showed that students were able to include all relevant aspects from the start. Criteria with clear benchmarks, such as requiring summaries to be 20-30% of the original text and limiting copied words to less than 30%, were quickly learned and adhered to by the students. However, the automated feedback did not effectively foster redundancy avoidance. The difficulty in avoiding redundancy can be understood through a combination of factors related to prior knowledge, the elaboration of mental models, and the design of the feedback itself. Activated and elaborated prior knowledge plays a critical role in how students process and organize information. When students possess insufficient or underdeveloped prior knowledge, it limits their ability to identify core ideas, make inferences, and effectively condense information (Kellogg, 1987; K. Kim et al., 2019; Mason et al., 2013; McCarthy & McNamara, 2021). This lack of knowledge often leads to surface-level processing, where students struggle to filter non-essential details, resulting in overly detailed or redundant summaries (Chi, 2009). In addition, students may have lacked elaborated mental models, which are crucial for restructuring and reorganizing text in a meaningful way. Summarizing requires integrating new information into the cognitive framework, and without well-developed mental models, students are likely to retain irrelevant details or repeat content, further contributing to redundancy (Hathorn & Rawson, 2012; M. K. Kim & McCarthy, 2021; van Dijk & Kintsch, 1983).

Another possible reason for the lack of redundancy avoidance can be attributed to the design of the automated feedback system. The feedback may not have been entirely clear in explaining why specific passages were marked as redundant. This lack of clarity may have prevented students from fully understanding and applying the feedback to improve their ability to write concisely. Moreover, the feedback primarily supported the revision phase of writing but did not address the planning phase, where students would need to engage more deeply in condensing and reorganizing the text's ideas (Chew et al., 2019). Without clear guidance during this important initial phase, students might have been unable to fully restructure the content into a concise format, leading to persistent redundancy. This issue was addressed in the third manuscript (see 3.1.5, 3.1.6). First, the formative provision of real-time feedback in several loops is examined by discussing the second research question:

3.1.2 Research Question 1.2: How does the formative nature of the automated feedback affect performance improvement?

Both the first and the third study positively associated the number of feedback loops to the text quality of the summaries. This suggests that the formative nature of the automated feedback fosters higher performance. According to the ITF model, the automated formative feedback facilitates an iterative interplay between the left and right feedback loops enabling a more structured and purposeful revision process. By aligning their internal reference standards with the external standard, students may not only have improved their current summaries but also internalized better summarizing skills for future tasks, leading to improved performance in subsequent sessions. External feedback in the right loop informed and adjusted the internal feedback in the left loop, which might have prompted students to continuously refine not only their summarizing skills but also self-regulation strategies (cf. Clark, 2012). By consistently highlighting areas for improvement, the automated formative feedback may have motivated students to maintain higher performance levels, thereby enhancing their motivational regulation. This is in line with former research that emphasizes that formative feedback fosters self-regulated learning (Black & Wiliam, 2018; Clark, 2012; Nicol & Macfarlane-Dick, 2006).

Furthermore, revising is not only about correcting but also about re-engaging with the material to deepen comprehension. When students are encouraged to revise their drafts, they may also restructure their ideas, identify gaps in their understanding, and incorporate new insights (cf. Bereiter & Scardamalia, 1987; Hayes et al., 1987). This iterative process fosters higher-order thinking and can significantly enhance the quality of the final summary. Moreover, students who frequently seek feedback may already possess, or be more likely to develop, more sophisticated strategies for processing and applying that feedback compared to their peers who do not engage with feedback as often (G. T. L. Brown et al., 2016; D. L. Butler & Winne, 2016; Narciss, 2017). They tend to approach feedback with an openness to change, demonstrating the ability to assess the relevance of external input, understand it, and adjust their internal standards accordingly. This is crucial, as feedback-seeking behavior is linked to deeper engagement with the material and a more deliberate revision process (Z. Zhang, 2020). The immediate availability of feedback after each revision may also serve as a direct incentive for students to uphold high standards and strive for continual improvement. Thus, automated formative feedback may motivate students to engage in careful revisions and promote the attainment of higher text quality, ultimately reinforcing the positive effects of interacting with automated feedback systems (M. Liu et al., 2017; Roscoe et al., 2015). Additionally, continuous feedback not only guides current work but may also be able to establish a routine for revising. This pattern is evident in the second study, where a cross-lagged panel analysis revealed consistent behavior in iterating feedback loops over time.

Yet, in both the first and the third studies, individual differences in the extent to which students engaged in multiple feedback loops were observed. Thus, in the following the effect of individual dispositions on the individual feedback engagement will be investigated and considerations how to operationalize the feedback engagement construct will be made.

3.1.3 Research Question 2.1: How consistent are students' cognitive, affective, and behavioral feedback engagement patterns across several time points?

The results of the second manuscript indicate that affective and behavioral feedback engagement are relatively stable across multiple sessions, with medium to high associations between engagement levels from one session to the next. This stability is consistent with theoretical understandings of the subconstructs presented before (cf. 1.3.2). Affective feedback engagement, which includes students' perceptions of feedback, the value they place on it, and the development of their educational identities, tends to develop gradually over time and is therefore less prone to major fluctuations within a semester (Chong, 2021; Sutton, 2012). Behavioral feedback engagement, which reflects the extent of student activity initiated by the feedback, also exhibits consistent patterns, with students maintaining stable engagement in feedback loops. This behavioral stability may indicate that students already possess routine strategies for processing feedback, as evidenced by their recurring engagement patterns (cf. 3.1.2, p. 46).

Yet, affective and behavioral feedback engagement are negatively correlated. This suggests that when students perceive feedback as less valuable, they tend to engage in more feedback loops. Contrary to conceptualizations of engagement with teacher or peer feedback, this increased behavioral engagement may act as a compensatory mechanism to reconcile doubts about the feedback's usefulness or informativeness. Students with a low affective engagement may still engage behaviorally by revisiting feedback multiple times, either to confirm their initial doubts or to test the feedback's validity. This behavior aligns with research indicating that students sometimes approach feedback with a skeptical or exploratory mindset, especially in settings where feedback is automated and might appear impersonal or unclear (Handley et al., 2011; Seifried et al., 2016; also cf. van der Kleij & Lipnevich, 2021). Furthermore, students may engage in iterative feedback loops as part of their self-regulation strategy to improve performance, even if they initially view the feedback skeptically (Zimmerman, 2000). In such cases, behavioral engagement may not stem from a positive affective reaction but rather from a pragmatic or instrumental approach, where students use feedback to refine their work regardless of their perceived value of it. This highlights a potential disconnect between affective engagement and behavioral engagement in automated feedback systems, particularly when students adopt a try-and-see strategy to gauge the feedback's impact on their performance.

In contrast, cognitive feedback engagement shows no consistency across sessions, indicating it may be more context-dependent and influenced by factors beyond prior cognitive engagement. This suggests that, while affective and behavioral engagement are shaped by more stable attitudes and habits, cognitive engagement either fluctuates based on task specific circumstances, such as text difficulty, or individual dispositions, such as the learners' performance levels (Hattie & Timperley, 2007; Shute, 2008). The second study found that prior performance positively impacts how feedback is interpreted and applied, subsequently influencing improvements in text quality. This finding suggests that higher-performing students may be more persistent in refining their summaries in response to feedback. This is also reflected in the moderate correlation between cognitive and behavioral feedback engagement confirming theory that suggests that cognitively engaged students know of the learning potential of the feedback and how to use it to improve their work (Price et al., 2011; Sutton, 2012). The correlation further supports the positive impact of iterative feedback loops on text quality, aligning with the findings of the first and third articles. Additionally, a slight positive correlation between cognitive and affective feedback engagement suggests that students who significantly improved their text quality place higher value on the automated formative feedback (Winstone et al., 2017). Overall, while students may approach automated formative feedback cautiously, the accessibility and motivating aspects of formative feedback encourage engagement in iterative revisions (M. Liu et al., 2017). This process may reduce the negative impact of affective feedback engagement in shaping learning outcomes, helping to bridge gaps in students' learning strategies and self-regulation abilities.

3.1.4 Research Question 2.2: How do the reciprocal relationships of cognitive and motivational resources impact feedback engagement processes in the perspective of time?

The second study used cross-lagged panel analysis across four intervention points, modeling the three dimensions of engagement as dependent variables, with cognitive and motivational resources as predictive variables. The complex model showed direct effects of 1) situational motivational subconstructs, and 2) mental effort on affective and behavioral feedback engagement, 3) two reinforcing reciprocal relationships, and 4) no effects of motivation and mental effort on cognitive feedback engagement. 1) The situational motivational subconstructs a) expectancy for success, b) intrinsic value and c) cost and their direct impact on feedback engagement are discussed. a) Expectancy for success explained a notable variation in affective and behavioral feedback engagement. Theory suggests that students who believe they are capable of succeeding are more likely to value feedback and act on it (Eccles & Wigfield, 2020). High expectancy for success typically encourages students to appreciate feedback and engage with it behaviorally by taking more iterations, as they see it as aligned with their goals (Putwain et al., 2019; Wu & Kang, 2021). However, in the second study of this dissertation, students with higher expectancy for success showed lower feedback acceptance, possibly because feedback inconsistent with their high expectations may have caused disappointment. Research on self-efficacy and feedback acceptance indicates that such students may protect their self-efficacy beliefs by questioning the feedback's accuracy or attributing poor performance to external factors like bad luck (Nease et al., 1999; Silver et al., 1995).

In contrast, b) intrinsic value, which represents the inherent interest and enjoyment a student finds in the task, impacted only affective feedback engagement. Thus, when students are intrinsically motivated, they are more likely to appreciate feedback and feel emotionally engaged with the learning process, as feedback aligns with their intrinsic enjoyment of the task (Durik et al., 2006; J. Guo et al., 2016; Watt et al., 2012). However, intrinsic value alone does not necessarily drive necessarily behavioral actions, such as revising and resubmitting work, which are more influenced by students' belief in their capability to succeed and the mental effort they have invested.

c) Cost did not directly impact any of the three dimensions of feedback engagement but did explain invested mental effort. According to cognitive load theory, high costs can make tasks feel burdensome, potentially reducing students' willingness to engage with feedback (Sweller, 2011). This effect was observed in the second study, where cost positively influenced students' perceptions of cognitive burden when processing the task and its feedback. However, in the context of a university tutorial, costs might have a limited impact on students' willingness to engage with feedback, as completing tasks is part of their academic responsibilities. Instead, students' perception of task burden and the effort required depends on specific costs, such as stress during processing or the need to forgo other activities to prioritize solving the task.

2) In turn, the perceived burden caused by the invested mental effort affected their feedback engagement. This is seen in the associations between invested mental effort and a) affective and b) behavioral feedback engagement. a) Students who invested more mental effort were more likely to question the valence of the feedback they received. This aligns with prior research on emotional responses to automated feedback, which highlights that students often perceive automated feedback as less credible or personalized compared to teacher-provided feedback (G. T. L. Brown et al., 2016; Seifried et al., 2016; van der Kleij & Lipnevich, 2021). When students experience a discrepancy between their perceived mental effort, their anticipated success, and the (potentially poorer) assessment provided by the feedback, they may react defensively or skeptically toward the feedback's validity and thus be inclined to disqualify or reinterpret the feedback rather than using it constructively for revision (Carless & Boud, 2018; Winstone et al., 2017). b) Higher mental effort resulted in higher behavioral feedback engagement. This relation demonstrates the consistency between students' perception of their investment of mental effort and their activity level with the automated formative feedback. Students who felt they dedicated more cognitive resources to improving their summaries also actually made more extensive use of the opportunities provided by the formative feedback.

Last, the findings also highlight the pivotal role of mental effort in the relationship between motivation and feedback engagement. Specifically, prior invested mental effort increases subsequent cost, which further amplifies mental effort (Feldon et al., 2019; Plass & Kalyuga, 2019).

3) The findings also revealed two reinforcing reciprocal relationships. The first is between expectancy for success and behavioral feedback engagement, where prior behavioral feedback engagement positively impacts subsequent expectancy, which further increases

behavioral feedback engagement. The second relationship is between intrinsic value and affective feedback engagement, with previous affective feedback engagement enhancing subsequent intrinsic value, which in turn boosts affective feedback engagement. This interconnected dynamic highlights the importance of leveraging these relationships to design interventions that promote sustained engagement and motivation. For instance, strategies aimed at encouraging multiple feedback loops and creating meaningful, relevant tasks can capitalize on these reciprocal effects to enhance both expectancy for success and intrinsic value over time. When students perceive a task as meaningful or believe in their ability to succeed, they are more intrinsically motivated to allocate greater mental effort. Enhancing these motivational drivers can thus foster deeper engagement with feedback and support the cognitive processes necessary for effective summarizing.

4) In contrast, none of the motivational sub-constructs nor invested mental effort directly impacted cognitive feedback engagement. Instead, cognitive feedback engagement—the change in text quality within a session—appears to rely more on prior knowledge than on motivational factors. This aligns with evidence from the second study, where previous performance predicted subsequent cognitive feedback engagement. Students who submitted higher quality summaries in the previous session were also more likely to make further improvements in the subsequent session, indicating higher cognitive feedback engagement. Interestingly, invested mental effort did not impact cognitive feedback engagement. However, cognitive feedback engagement influenced the level of invested mental effort, revealing a nuanced relationship between the two constructs. Specifically, students who successfully integrated feedback insights into their summaries and significantly improved their text quality also reported substantial mental effort. However, the perception of exerting significant mental effort

does not necessarily translate into successfully applying feedback insights or achieving measurable improvements in text quality.

In conclusion, the findings reveal that different dimensions of feedback engagement are uniquely shaped by students' motivational and cognitive resources, offering valuable insights into learner interactions with automated formative feedback in technology-enhanced environments. Encouraging students to allocate greater mental resources to seeking feedback more frequently emerges as a key strategy for enhancing behavioral engagement in the learning process. Building on these insights and considering the potential benefits of additional guidance during the planning phase as highlighted in research question 1.1 (p. 42), the next sections examine the support potential of instructional prompts.

3.1.5 Research Question 3.1: How can instructional prompts support the acquisition of summarizing skills compared to automated formative feedback?

Over the course of six sessions, students who received instructional prompts showed a noticeable improvement in their summarizing skills. This aligns with previous research demonstrating the positive impact of guided summarizing activities on enhancing students' summarizing abilities (Ahn, 2022; Lehmann et al., 2019; Wischgoll, 2017). However, while the group receiving prompts made progress, their improvement was less significant than that of the group receiving automated formative feedback.

In computer-based learning environments, instructional prompts play a crucial role in facilitating the interaction between students and the learning material, often acting as a bridge in the absence of real-time teacher intervention. The instructional prompts used in the third study were specifically designed to guide students through summarizing strategies. Summarizing strategies involve activities that coordinate both cognitive processes related to summarizing and self-regulation (Hidi & Anderson, 1986). For example, the prompts encouraged students to activate prior knowledge, ask clarifying questions, and identify key points in the text that were either unfamiliar or posed challenges. By prompting these activities, students were encouraged to align new information with their existing knowledge, integrate new concepts into their cognitive frameworks, and construct a coherent mental model which are key processes of summarizing (Doolittle et al., 2006; Palinscar & Brown, 1984). These steps are essential for deep engagement with the text and foster a more comprehensive understanding. As a result, this process not only aids comprehension but also supports the restructuring of the material, enabling students to summarize the content concisely in their own words (Garner, 1982; Head et al., 1989; Mok & Chan, 2016). Furthermore, the instructional prompts encouraged students to reflect on their understanding of the text and plan their summary. Berthold et al. (2007), Nückles et al. (2009), and Glogger et al. (2012) have shown that prompting a combination of cognitive and self-regulatory strategies is best for enhancing learning outcomes. Furthermore, repeated explicit prompting of these strategies may have helped students to become aware of and adopt these strategies leading to enhanced performance (Ahn, 2022; Ko, 2009). This is also shown in the third study where the experimental group who received instructional prompts improved their summaries throughout the intervention, maintaining elevated text quality in the post-test.

However, compared to the experimental group who received automated formative feedback, the improvement was relatively low. This may be explained with four reasons: First, while prompts promoted the planning phase of the writing process, they did not provide an external reference standard of the task criteria and a good summary. Therefore, the successful monitoring and evaluating of their summaries was dependent on the students' internal reference standard. The implications of relying solely on the internal reference standard to evaluate one's summary was discussed in the first research question (cf. 3.1.1). These implications also apply here. In contrast to the control group in the first study, which did not receive any support, the experimental group in the third study, however, was provided with prompts that promoted the use of summarizing strategies, and thus were able to improve slightly.

Second, to use prompts effectively, a common ground of understanding the purpose and aim of the prompts is necessary. However, varying levels of strategy sophistication and development of educational identities can influence students' interpretation of the prompts (cf. Sutton, 2012). Students with more advanced summarizing strategies or stronger metacognitive awareness are likely to recognize the deeper purpose of the prompts, using them as tools for reflection, cognitive engagement, or reinforcement of their existing approaches, seamlessly integrating them into their cognitive processes to enhance learning (cf. Engelmann et al., 2021). These students tend to see prompts as opportunities to refine their thinking, integrate new knowledge, and enhance their comprehension. In contrast, for high-performing students who have already cultivated autonomous and sophisticated learning strategies, prompts may sometimes be perceived as disruptive or redundant. Instead of enhancing their learning, the prompts may interfere with their established self-regulation processes, leading to frustration or disengagement (cf. Gidalevich & Kramarski, 2019; Kalyuga, 2014; Nückles et al., 2010). These students may feel that the prompts offer little added value, as they already possess the skills needed to self-regulate and summarize effectively without external guidance. On the other hand, students with less developed strategies or those still forming their educational identities may struggle to see the purpose of the prompts (cf. de Silva & Graham, 2015). Without a solid foundation in self-regulation, they may interpret the prompts as mere calls to action, rather than as opportunities for deeper cognitive engagement and reflection. Their limited understanding of the value behind these strategies can lead to a more superficial use of the prompts, where they focus on completing tasks rather than reflecting on the learning process.

Third, the prompts were provided only once at the beginning of the session, which limited their ability to offer ongoing support throughout the summarizing process. This single instance of guidance may not have supported students in managing the challenges that arise during revision, such as maintaining focus, or persisting through difficult sections of the text. Revising drafts is a critical component of summarizing because it allows students to refine their understanding, clarify their thoughts, and improve the organization and accuracy of their summaries (cf. Bereiter & Scardamalia, 1987; Hayes et al., 1987). As prompts did not address the iterative revision phase, students may have missed opportunities to critically review and improve their work, potentially leading to superficial or incomplete summaries.

Fourth, as highlighted in the second manuscript, motivational regulation is a key factor in the summarizing process. Additionally, as demonstrated in the first and third manuscripts, formative feedback helps foster motivation by encouraging students to engage more actively in feedback and revision cycles. Effective summarizing is a cognitively demanding task that requires sustained effort and engagement. Unlike the experimental groups in the first and third studies that received automated formative feedback, the group receiving prompts did not have motivational support throughout the session. As a result, these students may have struggled to maintain motivation and invest the effort needed to revise multiple times, especially when faced with challenging texts (cf. Gieseler et al., 2020; Kool et al., 2010; Shenhav et al., 2017; Yee & Braver, 2018). Consequently, their improvement was not as significant as that of the students who received automated formative feedback.

Another focus of the third manuscript was the question if the positive effects of providing instructional prompts and automated formative feedback can further promote the learning of effective summarizing skills when combined. Thus, the last research question is discussed in the following:

3.1.6 Research Question 3.2: How does a combination of instructional prompts and automated formative feedback affect the acquisition of summarizing skills?

Given the supportive potential of instructional prompts (Ahn, 2022; Wischgoll, 2017) and automated formative feedback (Lenhard et al., 2012; Wade-Stein & Kintsch, 2004) a combination of both should maximize learning potential. Thus, in the third study, an experimental group receiving both instructional prompts and automated formative feedback was compared to two other groups, one receiving only automated formative feedback and the other receiving only instructional prompts. While the combination group showed significantly greater improvement in summarizing skills compared to the prompts group, it did not outperform the feedback group over the course of the six sessions. The higher performance of both the combination and feedback groups, relative to the prompts group, can be attributed to the provision of automated formative feedback (see 3.1.1, 3.1.2, 3.1.5).

The combination of instructional prompts and automated formative feedback offers a nuanced approach to fostering summarizing skills, yet the expected synergy between the two does not necessarily yield higher learning outcomes compared to feedback alone. The primary driver behind the learning gains in both the feedback-only and combination groups appears to be the formative nature of the automated feedback, which effectively guided students through iterative cycles of revision. This feedback provided immediate, objective assessments of students' work, helping them align their drafts with the external standards of high-quality summaries. By continuously refining their texts based on this feedback, students in both groups improved their summarizing skills, demonstrating sustained progress over time (cf. 3.1.1, 3.1.2). The fact that the combination group did not outperform the feedback-only group suggests a potential ceiling effect, wherein the automated formative feedback already maximized students' learning potential. In other words, the feedback may have effectively elevated performance to a point where additional scaffolding, such as prompts, offered diminishing returns. Automated formative feedback not only pointed out areas for improvement but also promoted self-regulation by encouraging students to align their internal understanding of summary quality with external benchmarks. This alignment likely pushed students toward their upper limit of performance, leaving little room for the instructional prompts to further enhance learning outcomes.

The prompts, which aimed to activate cognitive processes such as prior knowledge and mental model construction, probably contributed to students' initial drafting. By guiding students to reflect on key aspects of the text and plan their summaries thoughtfully, the prompts facilitated the organization and structuring of ideas early in the writing process. However, this guidance may have led to a substitution effect, where the mental effort invested in following the prompts reduced the perceived need to fully engage with the automated feedback loops. The prompts may have been viewed as sufficient for addressing the task, causing students in the combination group to engage in fewer iterative revisions compared to those in the feedback-only group.

From an ITF model perspective, students in the feedback-only group may have approached the task with less structure initially but compensated for this by engaging more deeply with the feedback during the revision phase. The iterative nature of the feedback allowed them to progressively refine their drafts, aligning more closely with the quality standards over time. The external feedback loop acted as a formative mechanism, constantly informing their internal feedback processes, thereby promoting both the refinement of their summarizing skills and their ability to self-regulate. In contrast, the students in the combination group, may have seen the prompts as addressing most of the summarizing task requirements, potentially leading them to perceive the feedback as more of a summative evaluation than formative guidance. With the prompts already providing a scaffold for structuring their initial drafts, these students may not have felt the same need to engage in multiple feedback loops for further improvement, as demonstrated in the findings of the third article.

Table 4
Summary of the Key Findings of the Three Presented Manuscripts

	Article 1 **Effectiveness of Automated Formative Feedback**	**Article 2** **The Impact of Cognitive and Motivational Resources on Engagement with Automated Formative Feedback**	**Article 3** **The Interplay of Instructional Prompts and Automated Formative Feedback**
Aims	1.1) Investigating the effectiveness of automated formative feedback in improving summarizing skills. 1.2) Examining each aspect of summarizing (content, length, avoidance of copied words, redundancy avoidance) and their specific promotion by the automated formative feedback. 1.3) Exploring the impact of feedback loops on the improvement of summarizing skills.	2.1) Examining the development (indicated by time) of feedback engagement. 2.2) Investigating the reciprocal relationships of feedback engagement with cognitive and motivational resources	3.1) Investigating the effectiveness of instructional prompts versus automated formative feedback in improving summarizing skills. 3.2) Exploring if combining both prompts and feedback offers additional benefits in enhancing summarizing skills.
Methods	Pre-post-intervention study with six measurement points, experimental / control group design; $N = 183$ Ed. Elementary Education / M. A. Special Education students	Longitudinal study across six measurement points, cross-lagged-panel design, $N = 330$ B.Ed. Elementary Education students	Pre-post-intervention study with six measurement points and three group design, prompts-only, feedback-only, combination of prompts and feedback, $N = 254$
Analyses	Mixed-effects models	A priori power analysis with 90% power for detecting a medium effect, significance criterion of α =0.05, N =218; Cross-lagged panel analysis with SEM	Mixed-effects models
Findings	Automated formative feedback effectively fosters summarizing. Improvement in copied words avoidance, while including all important aspects and adhering to predefined length. Redundancy avoidance was not improved Number of feedback loops positively impacts summary text quality.	Previous affective and behavioral feedback (FB) engagement determines subsequent FB engagement. FB engagement depends on prior and concurrent cognitive and motivational resources. Low expectancy, low mental effort, and high interest increase affective FB engagement. Motivation reduces cost and mental effort and thus increases behavioral FB engagement.	Instructional prompts foster summarizing skills albeit to a lesser extent compared to automated formative feedback. A combination of feedback and prompts did not yield significant additional benefits. A substitution effect of prompts on feedback loops was observed.

3.2 Overall Contribution of the Dissertation to the Field

The studies and findings presented and discussed above have implications for 1) automated formative feedback in higher education, 2) the importance of learner factors, and 3) pedagogical considerations.

1) Automated feedback systems for summarizing have predominantly been employed in elementary schools, focusing on reading comprehension (Lenhard et al., 2012; Wade-Stein & Kintsch, 2004). However, developing such a system for the university context with the aim to additionally foster aspects of writing, such as expressing oneself in own words and redundancy avoidance, has not yet been explored. Furthermore, research in this context on explicit feedback on textual features and their long-term development has largely been overlooked. This research bridges that gap by adapting and enhancing a German feedback system (conText; Lenhard et al., 2013) originally designed for younger students and applying it to undergraduate university students. In the first manuscript, this feedback system proved effective in promoting summarizing skills, with students improving significantly in three of four evaluated criteria. This finding reinforces the potential for LSA-based feedback to be applicable beyond elementary and middle school settings, suggesting its versatility across educational levels.

The implications of this work extend beyond the development of summarizing skills. Automated formative feedback systems provide opportunities for individualized learning support without adding to educators' workloads. Through practicing effective summarizing, students acquire a foundational learning strategy that enhances their ability to engage with academic texts. Reading academic texts that address core aspects of teaching may encourage students to value scientific literature as a foundation for linking theory and practice in their ongoing professional development. This deeper engagement supports academic success, as effective learning strategies and a profound theoretical knowledge base are closely tied to improved performance in both academic studies (Dunlosky et al., 2013) and professional contexts (Kunina-Habenicht, 2020).

Furthermore, the results of the first and third manuscripts contribute to the field by systematically linking formative feedback to enhanced performance in summarizing tasks. While the theoretical connection between formative feedback and improved learning outcomes has long been established in educational research (Black & Wiliam, 2009, 2018; Clark, 2012; Graham, 2018), this dissertation provides empirical evidence supporting this relationship. This empirical validation underscores the practical efficacy of formative feedback, bridging the gap between theoretical assumptions and real-world application. By doing so, this work reinforces the importance of integrating formative feedback into educational contexts, providing a foundation for future research and practice to optimize their implementation and impact in automated systems.

2) The results of the second manuscript significantly advance the field of automated formative feedback research by illuminating the complex interplay between learner factors such as motivation and mental effort, and automated feedback engagement in educational contexts. The findings reveal that different forms of feedback engagement—cognitive, affective, and behavioral—respond uniquely to students' motivational and cognitive resources. This nuanced understanding deepens our knowledge of how learners interact with formative feedback in technology-enhanced environments, where ongoing engagement is shaped by individual differences in mental effort and motivation.

Moreover, the second manuscript contributes to theoretical discourse by addressing Eccles and Wigfield's (2020) call for more research on the evolution of subjective task value and competence belief hierarchies in response to feedback during specific tasks over varying timeframes. By investigating these processes, the manuscript begins to fill a critical gap in the literature, namely the scarcity of longitudinal studies exploring the reciprocal relationships between subjective task values, competence beliefs, mental effort, and automated feedback engagement.

While prior research has examined the isolated effects of cognitive load (Dong et al., 2020; Y. Liu & Sun, 2021) or motivation (Putwain et al., 2019; Sun & Rueda, 2012) on engagement, few studies have explored their interplay with automated feedback engagement or tracked these dynamics over extended periods (Han, 2017). This dissertation is one of the first to provide empirical evidence on these reciprocal relationships in a longitudinal design, thereby advancing both the theoretical understanding and practical application of formative feedback engagement in computer-based learning environments. These contributions lay the groundwork for future studies to further explore how students' feedback literacy needs to evolve to enable them to engage more effectively with automated formative feedback systems.

3) Results of the first and third manuscript deepened existing findings that summarizing skills can be supported through both instructional prompts and automated formative feedback, extending empirical evidence for the efficacy of these approaches (cf. Ahn, 2022; Chew et al., 2019; M. K. Kim & McCarthy, 2021; Wischgoll, 2017). Additionally, the third manuscript widened this understanding by comparing the separate effects of prompts and feedback to their combined use, offering novel insights into how these scaffolds complement each other. This represents a significant step toward exploring the supportive potential of different scaffolds and feedback mechanisms in computer-based learning environments, paving the way for future research to refine and optimize these strategies to cater to diverse learner needs.

Beyond its immediate contributions to summarizing skill development, this dissertation also has a distinct pedagogical dimension rooted in the dual-level pedagogical approach (Wahl, 2001). The learning environment implemented in the studies of this dissertation provided undergraduate elementary education students with an opportunity to improve their own summarizing skills while simultaneously reflecting on the pedagogical potential of the tool. This experience allows future teachers not only to engage with the tool as learners but also to consider its application in their own teaching practices. This dual-level approach fosters both subject-specific competencies and meta-pedagogical skills, promoting the integration of automated feedback systems into future classrooms. Research could further investigate whether early exposure to such systems in teacher education increases the likelihood of their adoption in teaching practice, contributing to broader educational innovation.

3.3 Critical Evaluation and Limitations of the Studies

In addition to the strengths outlined above, certain aspects of this dissertation also require a critical discussion. The limitations of the individual studies have already been discussed in detail in the respective manuscripts and are only partially mentioned here. Instead, this chapter focuses on aspects that go beyond the limitations discussed in the individual manuscripts. These critical aspects are primarily centered on three focal areas: 1) the samples, 2) the study setting and learning environment, and 3) the measures.

1) Regarding the samples, the three studies included in this dissertation were conducted with samples that provide a robust and reliable data foundation. Sample size plays a crucial role in determining the ability to detect hypothesized relationships and ensuring sufficient statistical power in the findings (Peers, 2006). However, the adequacy of a sample depends on various factors, including the study's purpose, expected effect sizes, the complexity of the statistical analyses, and the desired level of statistical power. Power analyses, described for instance in the second manuscript, demonstrated that the sample sizes used in these studies were appropriate for the conclusions drawn. Furthermore, the sample sizes were

sufficiently large to ensure the validity and reliability of the statistical methods employed. Yet, the findings are generalizable to a limited extent. All three samples consisted solely of undergraduate elementary education students from one university, and therefore are only representative in the context of this subpopulation. Additionally, although the studies were integrated into a university course, participation remained voluntary, which further constrained generalizability. Moreover, students who failed to complete all sessions or encountered technical issues were excluded, introducing the possibility of sample bias. Despite these limitations, the findings across all three studies align closely with existing research literature, showing consistency rather than contradiction.

2) Regarding the study setting and learning environment, in all three studies, a computer-based learning environment embedded within an online tutorial on summarizing was utilized. Students were required to complete assignments within a one- or two-week timeframe to fulfill course requirements, offering them flexibility in deciding when and where to engage with the tasks. This setup reflects the typical conditions under which online courses are implemented in universities, enhancing the ecological validity of the research. By mirroring real-world educational contexts, the findings are particularly relevant for understanding how students engage with online learning environments. However, this flexibility also introduced challenges in ensuring strict adherence to instructions and consistent, focused engagement with the program. To address these challenges, prompts were included to encourage students to work in a distraction-free environment and dedicate a 90-minute uninterrupted timeframe to the tasks. Despite these measures, the lack of direct control over the study environment raises concerns about internal validity. Variations in students' engagement levels or external distractions during task completion may have influenced the outcomes. Moreover, the flexibility provided may have led some students to approach the prompts and

feedback less seriously than they would in a more structured, supervised setting, potentially reducing the effectiveness of the interventions. To mitigate these concerns, future studies could incorporate measures such as activity logs or timed checkpoints to monitor and verify engagement, while still maintaining the ecological benefits of an online format. This approach would provide deeper insights into how students' surroundings influence outcomes and further strengthen both the internal and external validity of the findings.

Furthermore, one limitation of the interventions implemented in all studies is that the texts to be summarized were not counter-balanced, as they were presented in the same order across all three studies. This consistent order facilitated the comparison of differences between groups, as done in the first and third manuscripts, and allowed for inferences about influencing variables under uniform conditions, as explored in the second manuscript. However, this lack of counter-balancing introduces the possibility of a sequencing effect, where the order of text presentation influences participants' performance. For instance, the learning trajectories investigated in the first and third manuscripts could be impacted by this sequencing effect, as participants may respond differently to earlier versus later texts due to the order of presentation. The specific content or difficulty of the texts may also interact with their position in the sequence, potentially confounding the observed changes in performance over time. To address these concerns, future studies should randomize or counter-balance the order of texts to ensure a more robust investigation of the intervention's effects.

Another limitation is the missing assessment of students' prior knowledge of summarizing strategies, which could have influenced how they interacted with the feedback provided. Understanding students' prior summarizing strategies would have been particularly relevant for the feedback groups in the first and third studies, as it could have allowed to better evaluate the helpfulness of the feedback and prompts. This approach would have allowed for controlling prior knowledge not only through the pretest, which involved summarizing a text, but also through a more targeted evaluation of their existing summarizing strategies. Moreover, without measuring prior knowledge of summarizing strategies, it is difficult to account for the possibility of an expertise reversal effect in the prompts condition of the third manuscript. This effect occurs when students with a higher level of expertise in summarizing strategies find prompts unnecessary or redundant, potentially increasing extraneous cognitive load rather than supporting their learning (cf. Gidalevich & Kramarski, 2019; Kalyuga, 2014; Nückles et al., 2010). As a result, the prompts may have been less effective or even counterproductive for students who had already internalized effective summarizing strategies. Future research should incorporate measures of students' prior knowledge of summarizing strategies to better understand how both prompts and feedback interact with different levels of expertise and to refine these instructional tools to cater to diverse learner profiles.

Moreover, it was not directly assessed whether students learned the content of the texts they summarized. Summarizing is intended not only to produce concise representations of information but also to serve as a learning strategy that helps embed the text's content into students' cognitive schemas (Dunlosky et al., 2013). However, without a direct assessment of content retention or understanding, it remains unclear to what extent students internalized the material. Instead, the learning of the texts' content was inferred from the quality of the summaries, which is a valid approach since higher-quality summaries are expected to reflect deeper processing and integration of the texts' information, as outlined in the introduction (cf. p. 1f.). Nonetheless, future studies should include direct measures of content learning to provide a more comprehensive evaluation of summarizing as a learning strategy.

Last, in the third manuscript, the effects of prompts on the development of summarizing skills were examined and compared to those of automated feedback. However, a limitation of this study is that it did not assess whether students used the prompts in a meaningful way. Without knowing whether students used the prompts as intended, it's unclear if the observed improvements (or lack thereof) are due to the prompts' design or the students' personal engagement with them. Consequently, the impact of the prompts on the development of summarizing skills must be interpreted with caution. This limitation restricts the ability to evaluate the prompts' effectiveness beyond global inferences drawn from the improvement in summary quality and group comparisons.

3) With regard to the measures, all three studies used text quality (and changes in it) as the dependent variable. However, determining the quality of a text or summary is inherently challenging. This dissertation introduced an innovative approach to defining and operationalizing text quality, which represents both a strength and a limitation. While this approach provides a structured way to evaluate summaries, establishing a shared and universally accepted understanding of what constitutes text quality in this context remains complex and open to interpretation. To address this challenge, a formula was developed based on the works of Garner (1982), Head et al. (1989), and Sung et al. (2016), tailored to meet the specific needs of the three studies in this dissertation. This measure was validated through expert ratings to ensure its relevance and accuracy. While this method was carefully designed and represents a thoughtful and grounded approach, the measurement of text quality remains open to critique. Future research could benefit from dedicated studies aimed at further defining and refining the concept of text quality in summarizing, potentially offering a more robust and universally accepted measure. Nonetheless, the steps taken in this dissertation ensure that the approach is both methodologically sound and credible.

Another limitation lies in the measure of invested mental effort, as it does not clearly differentiate between types of cognitive load. Students' self-reports may reflect either effort willingly invested to master a meaningful and challenging task (germane cognitive load) or effort perceived as burdensome due to irrelevant demands (extraneous cognitive load). This distinction is critical because germane load supports deeper learning, while extraneous load can hinder learning outcomes and reduce engagement, particularly in automated feedback contexts (Plass et al., 2010; Sweller, 2020). As shown in the second manuscript, mental effort is also influenced by motivation, further complicating its interpretation. Intrinsic value encourages students to invest more effort due to interest and enjoyment, while perceived cost increases effort by making the task feel more strenuous. This dual influence means that mental effort can represent either productive engagement or the burden of overcoming challenges (Kalyuga & Singh, 2016; Leppink & Pérez-Fuster, 2019). The questions used to measure mental effort in the second study focused on understanding, processing, and summarizing the material, addressing students' attention, ease of paraphrasing, and summarizing effort. While this approach provides insight into cognitive processes, it does not fully distinguish whether reported effort stems from intrinsic, germane, or extraneous load. For example, higher perceived cost may increase effort allocated to attention and summarizing, while intrinsic value may drive effort through motivation and interest (Inzlicht et al., 2018). In summary, while the measure of mental effort captures relevant cognitive processes, its inability to differentiate between germane and extraneous cognitive loads is a limitation. Future research should incorporate refined measures to distinguish between productive and unproductive effort, offering deeper insights into the relationships between mental effort, motivation, and feedback engagement.

3.4 Outlook on Future Research

The following section provides an outlook on further questions raised by the approach and findings of this dissertation, focusing on four key topics: 1) Further automated feedback techniques and their comparison to this approach. 2) The importance of feedback literacy and promoting its development. 3) Student agency and its impact on instructional disobedience. 4) Further investigations about the complementary effect of prompts and feedback.

1) Since the inception of this dissertation project, there have been groundbreaking advancements in natural language processing, particularly with the development and public release of LLM, like GPT (OpenAI, 2022). These innovations have fundamentally transformed instructional practices and the ways in which assessments are conducted. For instance, students can now use ChatGPT to retrieve tailored information or even draft essays, while educators can create or customize ChatGPT prompts to provide students with automated feedback. However, feedback from LLMs is both human-like and inherently non-final or ambiguous, often requiring further interpretation by the user.

Future research could explore the conditions under which different types of automated feedback systems are most effective, particularly by comparing expert systems and LLMs. Expert systems, like the one employed in this dissertation, are designed to deliver structured, task-specific, and pedagogically aligned feedback. These systems are particularly effective in contexts with clear learning objectives and well-defined performance criteria, such as summarizing or writing tasks. However, their rigidity may limit their ability to address nuanced, open-ended needs that fall outside their programmed parameters. In contrast, LLMs, while more flexible and conversational, may lack the precision and reliability required for certain pedagogical applications. Understanding how these systems can complement each other or be optimized for specific educational scenarios would be a valuable direction for future research.

2) Feedback literacy, the ability to interpret, use, and act on feedback, becomes particularly important in the context of automated formative feedback systems, which often provide iterative suggestions that may not always feel entirely relevant or tailored to the individual. Students may perceive the feedback as misaligned with their expectations or needs, yet

valuable insights can still be drawn if feedback is approached with a flexible and reflective mindset. For instance, learning to interpret feedback beyond its immediate specificity can help students derive broader lessons for improvement. Developing these skills could empower students to leverage automated formative feedback systems more fully, transforming feedback into a dynamic tool for self-regulated learning. Combined with the findings of this dissertation on the interplay of cognitive and motivational factors in feedback engagement, future studies could explore how strengthening feedback literacy might enhance students' ability to navigate and benefit from automated formative feedback systems. This line of research could also examine how feedback literacy develops over time and its potential to foster deeper engagement, even when the feedback initially seems imperfect.

3) Expanding on the role of feedback literacy, another critical aspect to consider in automated feedback systems is the level of agency provided to learners. The degree of control and choice learners have in interacting with these systems can shape both the development of feedback literacy and the system's overall effectiveness in promoting meaningful learning outcomes. Highly structured environments offer clear guidance and scaffolded steps to support learning goals but may restrict opportunities for self-directed exploration and engagement. In contrast, open-ended environments provide flexibility and autonomy, which can foster creativity and deeper engagement, yet risk leaving some learners adrift without sufficient direction. This balance between structure and agency is particularly relevant in automated formative feedback systems. Systems that tightly control learning pathways may reduce the risk of instructional disobedience, where learners deviate from intended tasks or objectives, but may also stifle the opportunity for learners to engage with feedback in personally meaningful ways. Conversely, more open systems grant learners the freedom to interpret and apply feedback as they see fit, which can enhance agency but might also lead to inefficiencies or

misaligned learning outcomes if learners lack the skills to navigate such autonomy effectively. The discussion of agency in feedback systems raises important questions for future research and would be a valuable direction for future research.

4) Differentiating the roles of instructional prompts and automated formative feedback as it was investigated in the third manuscript has the potential to cater to individual differences in learning more effectively. Different learners may respond better to varying types of support; for instance, some students may benefit more from the structured guidance of instructional prompts that activate prior knowledge and help organize ideas, while others may find greater value in the iterative, targeted revisions facilitated by automated formative feedback. By offering multiple forms of scaffolding, learning environments can accommodate a broader range of students, allowing them to select the type of support that aligns with their cognitive needs and personal preferences. This flexibility might enable more individuals to engage meaningfully with the learning process, enhancing their ability to improve summarizing skills and self-regulation. To gain deeper insights into the interplay between scaffolds, future research could investigate how different learner prerequisites, such as varying levels of summarizing and self-regulative strategies, are associated with the use of prompts and feedback in automated formative feedback systems. Such studies could also explore whether and how providing prompts or feedback impacts learner behavior and engagement with the system. Addressing these questions would require research designs that focus on how learners with different skill levels and strategies interact with scaffolding conditions, and how these interactions affect their learning outcomes. This line of inquiry could reveal how instructional prompts and automated formative feedback can be tailored to meet diverse learner needs, reduce extraneous cognitive load, enhance germane cognitive load, and foster motivation and engagement.

3.5 Conclusion

This dissertation sought to investigate the effectiveness and conditions of automated formative feedback for fostering university students' summarizing skills, focusing on feedback characteristics, individual learner factors, and instructional factors. The evidence, gained in the three studies of this dissertation shows first that automated formative feedback effectively fosters summarizing skills in undergraduate students, especially when engaging in several feedback loops. Second, engagement depends on individual learner factors, such as motivation and mental effort, underlining the importance of encouraging students to allocate greater mental resources to seeking feedback more frequently. Third, with regard to instructional factors, adding instructional prompts does not enhance the effectiveness of automated formative feedback.

Consequently, with regard to the overall aim of the dissertation, it can be concluded, that automated formative feedback can be an important additional learning opportunity in higher education to foster necessary competencies for academic success without placing an excessive burden on teachers. However, in-line with the ITF model (Narciss, 2006, 2013, 2017), apart from the quality of an automated feedback system, individual learner factors and instructional factors decisively influence its effectiveness. Therefore, such systems not only have to be designed with regard to the immediate learning goal (i.e., summarizing), but teachers have to actively motivate students to engage with the system, for example by pointing out its usefulness and integrating automated approaches into the overall institutional educational strategies.

With these results, this dissertation contributes to closing significant research gaps in the field of automated formative feedback (see Chapter 1). First, this work is a pioneering effort to demonstrate the effectiveness of automated formative feedback in fostering summarizing skills among university students. Second, by presenting empirical evidence on the reciprocal relationships between subjective task values, competence beliefs, mental effort, and feedback engagement in a longitudinal context, light has now been shed on individual learners' cognitive and motivational factors. Third, the comparison of individual effects of instructional prompts and automated formative feedback with their combined use, contributes to the understanding of their interaction and complementary effects, which marks an important advancement in examining the supportive potential of various scaffolds and feedback mechanisms in computer-based learning environments. Overall, the insights provided in this dissertation, establish a foundation for future research on how feedback literacy can be developed to enhance student engagement with automated formative feedback systems. In doing so, this work offers a holistic approach to align automated innovations in education with the diverse needs of learners and the creation of learning opportunities.

Thus, the significance of this work lies in its ability to bridge the gap between technological advancements and the core principles of pedagogy. By addressing both the capabilities of automated formative feedback system and the needs of learners, this research contributes to creating more effective, inclusive, and impactful educational practices. In an era where technological disruptions also increasingly affect education, it is crucial to approach learning systems with a comprehensive perspective that balances innovation with pedagogical considerations.

While automated formative feedback systems offer immense potential to enhance learning processes, solely relying on them without considering the needs of learners and the broader educational context may result in suboptimal outcomes. Long-established pedagogical theories and principles must remain at the forefront of design and implementation to ensure that technology serves as a tool to support, rather than replace, meaningful educational interactions.

References

Acee, T. W., Weinstein, C. E., Hoang, T. V., & Flaggs, D. A. (2018). Value reappraisal as a conceptual model for task-value interventions. *The Journal of Experimental Education*, *86*(1), 69–85. https://doi.org/10.1080/00220973.2017.1381830

Ahn, S. (2022). Developing summary writing abilities of Korean EFL university students through teaching summarizing skills. *English Teaching*, *77*(2), 25–43. https://doi.org/10.15858/engtea.77.2.202206.25

Ali, N., Ahmed, L., & Rose, S. (2018). Identifying predictors of students' perception of and engagement with assessment feedback. *Active Learning in Higher Education*, *19*(3), 239–251. https://doi.org/10.1177/1469787417735609

Alipio, M. (2020). Predicting academic performance of college freshmen in the Philippines using psychological variables and expectancy-value beliefs to outcomes-based education: A path analysis. *Education & Administration*. https://doi.org/10.35542/osf.io/pra6z

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). Guildford Press.

Baadte, C., & Schnotz, W. (2014). Feedback effects on performance, motivation and mood: Are they moderated by the learner's self-cconcept? *Scandinavian Journal of Educational Research*, *58*(5), 570–591. https://doi.org/10.1080/00313831.2013.781059

Bangert-Drowns, R. L., Hurley, M. M., & Wilkinson, B. (2004). The effects of school-based writing-to-learn interventions on academic achievement: A meta-analysis. *Review of Educational Research*, *74*(1), 29–58.

Banihashem, S. K., Kerman, N. T., Noroozi, O., Moon, J., & Drachsler, H. (2024). Feedback sources in essay writing: Peer-generated or AI-generated feedback? *International Journal of Educational Technology in Higher Education*, *21*(1), 23. https://doi.org/10.1186/s41239-024-00455-4

Bannert, M. (2006). Effects of reflection prompts when learning with hypermedia. *Journal of Educational Computing Research*, *35*(4), 359–375. https://doi.org/10.2190/94V6-R58H-3367-G388

Barkela, V., Han, A., & Weber, A. M. (2024). Do student teachers experience self-worth threats in computational thinking? *Computers in Human Behavior Reports*, *15*, 100463. https://doi.org/10.1016/j.chbr.2024.100463

Barkela, V., & Leuchter, M. (2024a). Effectiveness of automated formative feedback in an online tutorial for promoting summarizing. *Journal of Educational Technology Development and Exchange*, *17*(1), 67–95. https://doi.org/10.18785/jetde.1701.04

Barkela, V., & Leuchter, M. (2024b). The interplay of instructional prompts and automated formative feedback in enhancing summarizing skills. *Manuscript Submitted for Publication*.

Barkela, V., Schmitt, L., & Leuchter, M. (2023). The impact of cognitive and motivational resources on engagement with automated formative feedback. *Contemporary Educational Psychology*, *75*, 102234. https://doi.org/10.1016/j.cedpsych.2023.102234

Becker-Mrotzek, M., Grabowski, J., Jost, J., Knopp, M., & Linnemann, M. (2014). Adressatenorientierung und Kohärenzherstellung im Text: Zum Zusammenhang kognitiver und sprachlich realisierter Teilkomponenten von Schreibkompetenz. *Halbjahresschrift Für Die Didaktik Der Deutschen Sprache Und Literatur*, *19*(37), 21–43.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? 🦜. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. https://doi.org/10.1145/3442188.3445922

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Routledge.

Berthold, K., Nückles, M., & Renkl, A. (2007). Do learning protocols support learning strategies and outcomes? The role of cognitive and metacognitive prompts. *Learning and Instruction*, *17*(5), 564–577. https://doi.org/10.1016/j.learninstruc.2007.09.007

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, *21*(1), 5–31. https://doi.org/10.1007/s11092-008-9068-5

Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, *25*(6), 551–575. https://doi.org/10.1080/0969594X.2018.1441807

Bong, M., Cho, C., Ahn, H. S., & Kim, H. J. (2012). Comparison of self-beliefs for predicting student motivation and achievement. *The Journal of Educational Research*, *105*(5), 336–352. https://doi.org/10.1080/00220671.2011.627401

Boud, D., & Molloy, E. (2013). Rethinking models of feedback for learning: The challenge of design. *Assessment & Evaluation in Higher Education*, *38*(6), 698–712. https://doi.org/10.1080/02602938.2012.691462

Brown, A. L., Day, J. D., & Jones, R. S. (1983). The development of plans for summarizing texts. *Child Developement*, *54*(4), 968–979.

Brown, G. T. L., Peterson, E. R., & Yao, E. S. (2016). Student conceptions of feedback: Impact on self-regulation, self-efficacy, and academic achievement. *The British Journal of Educational Psychology*, *86*(4), 606–629. https://doi.org/10.1111/bjep.12126

Butler, D. L., & Winne, P. H. (2016). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, *65*(3), 245–281. https://doi.org/10.3102/00346543065003245

Butler, J. A., & Britt, M. A. (2010). Investigating instruction for improving revision of argumentative essays. *Written Communication*, *28*(1), 70–96. https://doi.org/10.1177/0741088310387891

Canning, E. A., & Harackiewicz, J. M. (2019). Utility value and intervention framing. In K. A. Renninger & S. E. Hidi (Eds.), *The Cambridge handbook of motivation and learning* (pp. 645–662). Cambridge University Press.

Capa, R. L., & Audiffren, M. (2009). How does achievement motivation influence mental effort mobilization? Physiological evidence of deteriorative effects of negative affects on the level of engagement. *International Journal of Psychophysiology*, *74*(3), 236–242. https://doi.org/10.1016/j.ijpsycho.2009.09.007

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, *43*(8), 1315–1325. https://doi.org/10.1080/02602938.2018.1463354

Carless, D., & Winstone, N. (2023). Teacher feedback literacy and its interplay with student feedback literacy. *Teaching in Higher Education*, *28*(1), 150–163. https://doi.org/10.1080/13562517.2020.1782372

Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., Chen, H., Yi, X., Wang, C., Wang, Y., Ye, W., Zhang, Y., Chang, Y., Yu, P. S., Yang, Q., & Xie, X. (2024). A survey on evaluation of large language models. *ACM Transactions on Intelligent Systems and Technology*, *15*(3), 1–45. https://doi.org/10.1145/3641289

Chen, C.-M., & Huang, S.-H. (2014). Web-based reading annotation system with an attention-based self-regulated learning mechanism for promoting reading performance: Attention-based self-regulated learning mechanism. *British Journal of Educational Technology*, *45*(5), 959–980. https://doi.org/10.1111/bjet.12119

Cheng, X., & Zhang, L. J. (2024). Engaging secondary school students with peer feedback in L2 writing classrooms: A mixed-methods study. *Studies in Educational Evaluation*, *81*, 101337. https://doi.org/10.1016/j.stueduc.2024.101337

Chew, C. S., Idris, N., Loh, E. F., Wu, W. V., Chua, Y. P., & Bimba, A. T. (2019). The effects of a theory-based summary writing tool on students' summary writing. *Journal of Computer Assisted Learning*, *35*(3), 435–449. https://doi.org/10.1111/jcal.12349

Chi, M. T. H. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, *1*(1), 73–105. https://doi.org/10.1111/j.1756-8765.2008.01005.x

Chong, S. W. (2021). Reconsidering student feedback literacy from an ecological perspective. *Assessment & Evaluation in Higher Education*, *46*(1), 92–104. https://doi.org/10.1080/02602938.2020.1730765

Clark, I. (2012). Formative assessment: Assessment is for self-regulated learning. *Educational Psychology Review*, *24*(2), 205–249. https://doi.org/10.1007/s10648-011-9191-6

Conroy, D. E. (2004). The unique psychological meanings of multidimensional fears of failing. *Journal of Sport and Exercise Psychology*, *26*(3), 484–491. https://doi.org/10.1123/jsep.26.3.484

de Araujo Guerra Grangeia, T., de Jorge, B., Franci, D., Martins Santos, T., Vellutini Setubal, M. S., Schweller, M., & de Carvalho-Filho, M. A. (2016). Cognitive load and self-determination theories applied to E-learning: Impact on students' participation and academic performance. *PLOS ONE*, *11*(3), e0152462. https://doi.org/10.1371/journal.pone.0152462

de Silva, R., & Graham, S. (2015). The effects of strategy instruction on writing strategy use for students of different proficiency levels. *System*, *53*, 47–59. https://doi.org/10.1016/j.system.2015.06.009

DeBacker, T. K., & Nelson, R. M. (1999). Variations on an expectancy-value model of motivation in science. *Contemporary Educational Psychology*, *24*(2), 71–94.

Deeva, G., Bogdanova, D., Serral, E., Snoeck, M., & De Weerdt, J. (2021). A review of automated feedback systems for learners: Classification framework, challenges, and opportunities. *Computers & Education*, *162*, 104094. https://doi.org/10.1016/j.compedu.2020.104094

Delen, E., Liew, J., & Willson, V. (2014). Effects of interactivity and instructional scaffolding on learning: Self-regulation in online video-based environments. *Computers & Education*, *78*, 312–320. https://doi.org/10.1016/j.compedu.2014.06.018

Dietrich, J., Moeller, J., Guo, J., Viljaranta, J., & Kracke, B. (2019). In-the-moment profiles of expectancies, task values, and costs. *Frontiers in Psychology*, *10*, 1662. https://doi.org/10.3389/fpsyg.2019.01662

Dietrich, J., Viljaranta, J., Moeller, J., & Kracke, B. (2017). Situational expectancies and task values: Associations with students' effort. *Learning and Instruction*, *47*, 53–64. https://doi.org/10.1016/j.learninstruc.2016.10.009

Dole, J. A., Duffy, G. G., Roehler, L. R., & Pearson, P. D. (1991). Moving from the old to the new: Research on reading comprehension instruction. *Review of Educational Research*, *61*(2), 239–264.

Doménech-Betoret, F., Abellán-Roselló, L., & Gómez-Artiga, A. (2017). Self-efficacy, satisfaction, and academic achievement: The mediator role of students' expectancy-value beliefs. *Frontiers in Psychology*, *8*, 1193. https://doi.org/10.3389/fpsyg.2017.01193

Dong, A., Jong, M. S.-Y., & King, R. B. (2020). How does prior knowledge influence learning engagement? The mediating roles of cognitive load and help-seeking. *Frontiers in Psychology*, *11*, 591203. https://doi.org/10.3389/fpsyg.2020.591203

Doolittle, P. E., Hicks, D., Triplett, C. F., Nichols, W. D., & Young, C. A. (2006). Reciprocal teaching for reading comprehension in higher education: A strategy for fostering the deeper understanding of texts. *International Journal of Teaching and Learning in Higher Education*, *17*(2), 106–118.

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

Durik, A. M., Vida, M., & Eccles, J. S. (2006). Task values and ability beliefs as predictors of high school literacy choices: A developmental analysis. *Journal of Educational Psychology*, *98*(2), 382–393. https://doi.org/10.1037/0022-0663.98.2.382

Eccles, J. S. (2016). Engagement: Where to next? *Learning and Instruction*, *43*, 71–75. https://doi.org/10.1016/j.learninstruc.2016.02.003

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

Elliot, A. J., & Church, M. A. (1997). A hierarchical model of approach and avoidance achievement motivation. *Journal of Personality and Social Psychology*, *72*(1), 218–232. https://doi.org/10.1037/0022-3514.72.1.218

Ellis, R. (2010). Epilogue: A framework for investigating oral and written corrective feedback. *Studies in Second Language Acquisition*, *32*(2), 335–349. https://doi.org/10.1017/S0272263109990544

Engelmann, K., & Bannert, M. (2021). Analyzing temporal data for understanding the learning process induced by metacognitive prompts. *Learning and Instruction*, *72*, 101205. https://doi.org/10.1016/j.learninstruc.2019.05.002

Engelmann, K., Bannert, M., & Melzner, N. (2021). Do self-created metacognitive prompts promote short- and long-term effects in computer-based learning environments? *Research and Practice in Technology Enhanced Learning*, *16*(1), 1–21. https://doi.org/10.1186/s41039-021-00148-w

Evans, C. (2013). Making sense of assessment feedback in higher education. *Review of Educational Research*, *83*(1), 70–120. https://doi.org/10.3102/0034654312474350

Fan, W., & Williams, C. M. (2010). The effects of parental involvement on students' academic self-efficacy, engagement and intrinsic motivation. *Educational Psychology*, *30*(1), 53–74. https://doi.org/10.1080/01443410903353302

Fan, Y., & Xu, J. (2020). Exploring student engagement with peer feedback on L2 writing. *Journal of Second Language Writing*, *50*, 100775. https://doi.org/10.1016/j.jslw.2020.100775

Feldon, D. F., Callan, G., Juth, S., & Jeong, S. (2019). Cognitive load as motivational cost. *Educational Psychology Review*, *31*(2), 319–337. https://doi.org/10.1007/s10648-019-09464-6

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

Flake, J. K., Barron, K. E., Hulleman, C., McCoach, B. D., & Welsh, M. E. (2015). Measuring cost: The forgotten component of expectancy-value theory. *Contemporary Educational Psychology*, *41*, 232–244. https://doi.org/10.1016/j.cedpsych.2015.03.002

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, *74*(1), 59–109. https://doi.org/10.3102/00346543074001059

Friend, R. (2001). Effects of strategy instruction on summary writing of college students. *Contemporary Educational Psychology*, *26*(1), 3–24. https://doi.org/10.1006/ceps.1999.1022

Frömer, R., Lin, H., Dean Wolf, C. K., Inzlicht, M., & Shenhav, A. (2021). Expectations of reward and efficacy guide cognitive control allocation. *Nature Communications*, *12*(1), 1030. https://doi.org/10.1038/s41467-021-21315-z

Gan, Mark J. S., & Hattie, J. (2014). Prompting secondary students' use of criteria, feedback specificity and feedback levels during an investigative task. *Instructional Science*, *42*, 861–878. https://doi.org/10.1007/s11251-014-9319-4

Garner, R. (1982). Efficient text summarization costs and benefits. *The Journal of Educational Research*, *75*(5), 275–279. https://doi.org/10.1080/00220671.1982.10885394

Gaspard, H., Dicke, A.-L., Flunger, B., Brisson, B. M., Häfner, I., Nagengast, B., & Trautwein, U. (2015). Fostering adolescents' value beliefs for mathematics with a relevance intervention in the classroom. *Developmental Psychology*, *51*(9), 1226–1240. https://doi.org/10.1037/dev0000028

Gentner, N., & Seufert, T. (2020). The double-edged interactions of prompts and self-efficacy. *Metacognition and Learning*, *15*(2), 261–289. https://doi.org/10.1007/s11409-020-09227-7

Gidalevich, S., & Kramarski, B. (2019). The value of fixed versus faded self-regulatory scaffolds on fourth graders' mathematical problem solving. *Instructional Science*, *47*(1), 39–68. https://doi.org/10.1007/s11251-018-9475-z

Gieseler, K., Inzlicht, M., & Friese, M. (2020). Do people avoid mental effort after facing a highly demanding task? *Journal of Experimental Social Psychology*, *90*, 104008. https://doi.org/10.1016/j.jesp.2020.104008

Glogger, I., Schwonke, R., Holzäpfel, L., Nückles, M., & Renkl, A. (2012). Learning strategies assessed by journal writing: Prediction of learning outcomes by quantity, quality, and combinations of learning strategies. *Journal of Educational Psychology*, *104*(2), 452–468. https://doi.org/10.1037/a0026683

Graham, S. (2006). Strategy instruction and the teaching of writing: A meta-analysis. In C. A. McArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 187–207). The Guilford Press.

Graham, S. (2018). Instructional feedback in writing. In A. A. Lipnevich & J. K. Smith (Eds.), *The Cambridge handbook of instructional feedback* (1st ed., pp. 145–168). Cambridge University Press. https://doi.org/10.1017/9781316832134.009

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

Green, J. M., & Holman, J. (2021). Cultivating the strategy of summarizing sequentiale expository text: Scaffolds and supports for the intermediate grades. *Literacy Practice and Research*, *46*(1). https://doi.org/10.25148/lpr.009343

Guo, J., Nagengast, B., Marsh, H. W., Kelava, A., Gaspard, H., Brandt, H., Cambria, J., Flunger, B., Dicke, A.-L., Häfner, I., Brisson, B., & Trautwein, U. (2016). Probing the unique contributions of self-concept, task values, and their interactions using multiple value facets and multiple academic outcomes. *AERA Open*, *2*(1). https://doi.org/10.1177/2332858415626884

Guo, Q., Feng, R., & Hua, Y. (2022). How effectively can EFL students use automated written corrective feedback (AWCF) in research writing? *Computer Assisted Language Learning*, *35*(9), 2312–2331. https://doi.org/10.1080/09588221.2021.1879161

Gurlitt, J., & Renkl, A. (2010). Prior knowledge activation: How different concept mapping tasks lead to substantial differences in cognitive processes, learning outcomes, and perceived self-efficacy. *Instructional Science*, *38*(4), 417–433. https://doi.org/10.1007/s11251-008-9090-5

Han, Y. (2017). Mediating and being mediated: Learner beliefs and learner engagement with written corrective feedback. *System*, *69*, 133–142. https://doi.org/10.1016/j.system.2017.07.003

Han, Y. (2019). Written corrective feedback from an ecological perspective: The interaction between the context and individual learners. *System*, *80*, 288–303. https://doi.org/10.1016/j.system.2018.12.009

Han, Y., & Hyland, F. (2015). Exploring learner engagement with written corrective feedback in a Chinese tertiary EFL classroom. *Journal of Second Language Writing*, *30*, 31–44. https://doi.org/10.1016/j.jslw.2015.08.002

Han, Y., & Xu, Y. (2021). Student feedback literacy and engagement with feedback: A case study of Chinese undergraduate students. *Teaching in Higher Education*, *26*(2), 181–196. https://doi.org/10.1080/13562517.2019.1648410

Handley, K., Price, M., & Millar, J. (2011). Beyond ‘doing time’: Investigating the concept of student engagement with feedback. *Oxford Review of Education*, *37*(4), 543–560. https://doi.org/10.1080/03054985.2011.604951

Hathorn, L. G., & Rawson, K. A. (2012). The roles of embedded monitoring requests and questions in improving mental models of computer-based scientific text. *Computers & Education*, *59*(3), 1021–1031. https://doi.org/10.1016/j.compedu.2012.04.014

Hattie, J., Biggs, J., & Purdie, N. (1996). Effects of learning skills interventions on student learning: A meta-analysis. *Review of Educational Research*, *66*(2), 99–136. https://doi.org/10.3102/00346543066002099

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487

Hayes, J. R., Flower, L. S., Shriver, K. A., Stratman, J. F., & Carey, L. (1987). Cognitive processes in revision. In S. Rosenberg (Ed.), *Advances in applied psycholinguistics: Vol. II. Reading, writing and language learning* (Vol. 2, pp. 176–240). Cambridge Univ. Pr.

Head, M. H., Readence, J. E., & Buss, R. R. (1989). An examination of summary writing as a measure of reading comprehension. *Reading Research and Instruction*, *28*(4), 1–11. https://doi.org/10.1080/19388078909557982

Hecht, C. A., Grande, M. R., & Harackiewicz, J. M. (2021). The role of utility value in promoting interest development. *Motivation Science*, *7*(1), 1–20. https://doi.org/10.1037/mot0000182

Hefter, M. H., Kubik, V., & Berthold, K. (2023). Can prompts improve self-explaining an online video lecture? Yes, but do not disturb! *International Journal of Educational Technology in Higher Education*, *20*(1), 15. https://doi.org/10.1186/s41239-023-00383-9

Hidi, S., & Anderson, V. (1986). Producing written summaries: Task demands, cognitive operations, and implications for instruction. *Review of Educational Research*, *56*(4), 473–493. https://doi.org/10.3102/00346543056004473

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist*, *41*(2), 111–127. https://doi.org/10.1207/s15326985ep4102_4

Hill, M. (1991). Writing summaries promotes thinking and learning across the curriculum—But why are they so difficult to write? *Journal of Reading*, *34*(7), 536–539. http://www.jstor.org/stable/40014578

Hulleman, C. S., Godes, O., Hendricks, B. L., & Harackiewicz, J. M. (2010). Enhancing interest and performance with a utility value intervention. *Journal of Educational Psychology*, *102*(4), 880–895. https://doi.org/10.1037/a0019506

Hulleman, C. S., & Harackiewicz, J. M. (2020). The utility value intervention. In Walton, Gregory M. & Crum, Alia J. (Eds.), *Handbook of wise interventions: How social-psychological insights can help solve problems* (pp. 100–125). Guildford Press.

Hyland, K., & Zhang, Z. V. (2018). Student engagement with teacher and automated feedback on L2 writing. *Assessing Writing*, *36*, 90–102. https://doi.org/10.1016/j.asw.2018.02.004

Inzlicht, M., Shenhav, A., & Olivola, C. Y. (2018). The effort paradox: Effort Is both costly and valued. *Trends in Cognitive Sciences*, *22*(4), 337–349. https://doi.org/10.1016/j.tics.2018.01.007

Iraj, H., Fudge, A., Khan, H., Faulkner, M., Pardo, A., & Kovanović, V. (2021). Narrowing the feedback gap: Examining student engagement with personalized and actionable feedback messages. *Journal of Learning Analytics*, *8*(3), 101–116. https://doi.org/10.18608/jla.2021.7184

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

Johnson, A. M., Azevedo, R., & D'Mello, S. K. (2011). The temporal and dynamic nature of self-regulatory processes during independent and externally assisted hypermedia learning. *Cognition and Instruction*, *29*(4), 471–504. https://doi.org/10.1080/07370008.2011.610244

Johnson, M. L., & Sinatra, G. M. (2013). Use of task-value instructional inductions for facilitating engagement and conceptual change. *Contemporary Educational Psychology*, *38*(1), 51–63. https://doi.org/10.1016/j.cedpsych.2012.09.003

Jung, Y., & Lee, J. (2018). Learning engagement and persistence in massive open online courses (MOOCS). *Computers & Education*, *122*, 9–22. https://doi.org/10.1016/j.compedu.2018.02.013

Kabudi, T., Pappas, I., & Olsen, D. H. (2021). AI-enabled adaptive learning systems: A systematic mapping of the literature. *Computers and Education: Artificial Intelligence*, *2*, 100017. https://doi.org/10.1016/j.caeai.2021.100017

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

Kalyuga, S., & Singh, A.-M. (2016). Rethinking the boundaries of cognitive load theory in complex learning. *Educational Psychology Review*, *28*(4), 831–852. https://doi.org/10.1007/s10648-015-9352-0

Keck, C. (2006). The use of paraphrase in summary writing: A comparison of L1 and L2 writers. *Journal of Second Language Writing*, *15*(4), 261–278. https://doi.org/10.1016/j.jslw.2006.09.006

Kellogg, R. T. (1987). Effects of topic knowledge on the allocation of processing time and cognitive effort to writing processes. *Memory & Cognition*, *15*(3), 256–266. https://doi.org/10.3758/BF03197724

Kellogg, R. T. (1988). Attentional overload and writing performance: Effects of rough draft and outline strategies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(2), 355–365. https://doi.org/10.1037/0278-7393.14.2.355

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. (2012). The development of writing expertise. In E. L. Grigorenko, E. Mambrino, & D. D. Preiss (Eds.), *Writing: A mosaic of new perspectives* (pp. 109–124). Psychology Press.

Kim, K., Clarianay, R. B., & Kim, Y. (2019). Automatic representation of knowledge structure: Enhancing learning through knowledge structure reflection in an online course. *Educational Technology Research and Development*, *67*(1), 105–122. https://doi.org/10.1007/s11423-018-9626-6

Kim, M. K., & McCarthy, K. S. (2021). Improving summary writing through formative feedback in a technology-enhanced learning environment. *Journal of Computer Assisted Learning*, *37*(3), 684–704. https://doi.org/10.1111/jcal.12516

Kim, Y.-S. G. (2020). Hierarchical and dynamic relations of language and cognitive skills to reading comprehension: Testing the direct and indirect effects model of reading (DIER). *Journal of Educational Psychology*, *112*(4), 667–684. https://doi.org/10.1037/edu0000407

Kinsey, A. W. (2022). *The relationship of cognitive load and motivation as a predictor of persistence in accelerated online asynchronous courses* [Doctoral dissertation, University of Memphis]. https://digitalcommons.memphis.edu/cgi/viewcontent.cgi?article=4538&context=etd

Kintsch, E. (1990). Macroprocesses and microprocesses in the development of summarization skill. *Cognition and Instruction*, *7*(3), 161–195. https://doi.org/10.1207/s1532690xci0703_1

Kintsch, E., Steinhart, D., Stahl, G., LSA Research Group, Matthews, C., & Lamb, R. (2000). Developing summarization skills through the use of LSA-based feedback. *Interactive Learning Environments*, *8*(2), 87–109. https://doi.org/10.1076/1049-4820(200008)8:2;1-B;FT087

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*(5), 363–393.

Kirkland, M. R., & Saunders, M. A. P. (1991). Maximizing student performance in summary writing: Managing cognitive load. *TESOL Quarterly*, *25*(1), 105–121. https://doi.org/10.2307/3587030

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

Ko, M. H. (2009). Summary writing instruction and student learning outcomes. *English Teaching*, *2*(64), 125–149.

Koltovskaia, S., & Mahapatra, S. (2022). Student engagement with computer-mediated teacher written corrective feedback: A case study. *The JALT CALL Journal*, *18*(2), 286–315. https://doi.org/10.29140/jaltcall.v18n2.519

Kool, W., McGuire, J. T., Rosen, Z. B., & Botvinick, M. M. (2010). Decision making and the avoidance of cognitive demand. *Journal of Experimental Psychology: General*, *139*(4), 665–682. https://doi.org/10.1037/a0020198

Kool, W., Shenhav, A., & Botvinick, M. M. (2017). Cognitive control as cost-benefit decision making. In T. Egner (Ed.), *The Wiley handbook of cognitive control* (pp. 167–189). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118920497.ch10

Krause, U.-M., & Stark, R. (2010). Reflection in example- and problem-based learning: Effects of reflection prompts, feedback and cooperative learning. *Evaluation & Research in Education*, *23*(4), 255–272. https://doi.org/10.1080/09500790.2010.519024

Kunina-Habenicht, O. (2020). Wissen ist Macht: Ein Plädoyer für ein wissenschaftliches Lehramtsstudium. In C. Scheid & T. Wenzl (Eds.), *Wieviel Wissenschaft braucht die Lehrerbildung?* (pp. 109–126). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-23244-3_6

Landauer, T. K. (1999). Latent semantic analysis: A theory of the psychology of language and mind. *Discourse Processes*, *27*(3), 303–310. https://doi.org/10.1080/01638539909545065

Lee, C. (2020). A study of adolescent English learners' cognitive engagement in writing while using an automated content feedback system. *Computer Assisted Language Learning*, *33*(1–2), Article 1–2. https://doi.org/10.1080/09588221.2018.1544152

Lee, J., Lee, M., & Bong, M. (2013). High value with low perceived competence as an amplifier of self-worth threat. In D. M. McInerney, H. W. Marsh, R. G. Craven, & F. Guay (Eds.), *Theory driving research: New wave perspectives on self-processed and human development* (pp. 205–231). IAP Information Age Publishing.

Lehmann, T., Hähnlein, I., & Ifenthaler, D. (2014). Cognitive, metacognitive and motivational perspectives on preflection in self-regulated online learning. *Computers in Human Behavior*, *32*, 313–323. https://doi.org/10.1016/j.chb.2013.07.051

Lehmann, T., Rott, B., & Schmidt-Borcherding, F. (2019). Promoting pre-service teachers' integration of professional knowledge: Effects of writing tasks and prompts on learning from multiple documents. *Instructional Science*, *47*(1), 99–126. https://doi.org/10.1007/s11251-018-9472-2

Lenhard, W. (2008). *Bridging the gap to natural language: A review on intelligent tutoring systems based on Latent Semantic Analysis*. https://opus.bibliothek.uni-wuerzburg.de/files/2397/Lenhard_Bridging_the_Gap.pdf

Lenhard, W., Baier, H., Endlich, D., Lenhard, A., Schneider, W., & Hoffmann, J. (2012). Computerunterstützte Leseverständnisförderung: Die Effekte automatisch generierter Rückmeldungen. *Zeitschrift Für Pädagogische Psychologie*, *26*(2), 135–148. https://doi.org/10.1024/1010-0652/a000066

Lenhard, W., Baier, H., Lenhard, A., Hoffmann, J., & Schneider, W. (2013). *ConText: Förderung des Leseverständnisses durch das Arbeiten mit Texten: Manual*. Hogrefe.

Lenhard, W., & Lenhard, A. (2014). *Berechnung des Lesbarkeitsindex LIX nach Björnson*. https://doi.org/10.13140/RG.2.1.1512.3447

Leopold, C., Brückner, A., & Dutke, S. (2019). Summarizing as a strategy for science text comprehension: Text-based versus content-based processing. *Discourse Processes*, *56*(8), 728–747. https://doi.org/10.1080/0163853X.2018.1563849

Leppink, J., & Pérez-Fuster, P. (2019). Mental effort, workload, time on task, and certainty: Beyond linear models. *Educational Psychology Review*, *31*(2), 421–438. https://doi.org/10.1007/s10648-018-09460-2

Li, H., Cai, Z., & Graesser, A. C. (2018). Computerized summary scoring: Crowdsourcing-based latent semantic analysis. *Behavior Research Methods*, *50*(5), 2144–2161. https://doi.org/10.3758/s13428-017-0982-7

Li, J. (2014). The role of reading and writing in summarization as an integrated task. *Language Testing in Asia*, *4*(1), 3. https://doi.org/10.1186/2229-0443-4-3

Lim, L., Bannert, M., van der Graaf, J., Singh, S., Fan, Y., Surendrannair, S., Rakovic, M., Molenaar, I., Moore, J., & Gašević, D. (2023). Effects of real-time analytics-based personalized scaffolds on students' self-regulated learning. *Computers in Human Behavior*, *139*, 107547. https://doi.org/10.1016/j.chb.2022.107547

Linnenbrink-Garcia, L., Perez, T., Barger, M. M., Wormington, S. V., Godin, E., Snyder, K. E., Robinson, K., Sarkar, A., Richman, L. S., & Schwartz-Bloom, R. (2018). Repairing the leaky pipeline: A motivationally supportive intervention to enhance persistence in undergraduate science pathways. *Contemporary Educational Psychology*, *53*, 181–195. https://doi.org/10.1016/j.cedpsych.2018.03.001

Lipnevich, A. A., & Panadero, E. (2021). A review of feedback models and theories: Descriptions, definitions, and conclusions. *Frontiers in Education*, *6*, 720195. https://doi.org/10.3389/feduc.2021.720195

Lipnevich, A. A., & Smith, J. K. (2009a). Effects of differential feedback on students' examination performance. *Journal of Experimental Psychology: Applied*, *15*(4), 319–333. https://doi.org/10.1037/a0017841

Lipnevich, A. A., & Smith, J. K. (2009b). "I really need feedback to learn": Students' perspectives on the effectiveness of the differential feedback messages. *Educational Assessment, Evaluation and Accountability*, *21*(4), 347–367. https://doi.org/10.1007/s11092-009-9082-2

Liu, M., Li, Y., Xu, W., & Liu, L. (2017). Automated essay feedback generation and its impact on revision. *IEEE Transactions on Learning Technologies*, *10*(4), 502–513. https://doi.org/10.1109/tlt.2016.2612659

Liu, Y., & Sun, J. C.-Y. (2021). The mediation effects of task strategies on the relationship between engagement and cognitive load in a smart instant feedback system. *2021 International Conference on Advanced Learning Technologies (ICALT)*, 195–199. https://doi.org/10.1109/ICALT52272.2021.00065

Lyu, B., & Lai, C. (2022). Analysing learner engagement with native speaker feedback on an educational social networking site: An ecological perspective. *Computer Assisted*

Language Learning, *37*(1–2), 114–148. https://doi.org/10.1080/09588221.2022.2030364

Maier, U., & Klotz, C. (2022). Personalized feedback in digital learning environments: Classification framework and literature review. *Computers and Education: Artificial Intelligence*, *3*, 100080. https://doi.org/10.1016/j.caeai.2022.100080

Malecka, B., & Boud, D. (2021). Fostering student motivation and engagement with feedback through ipsative processes. *Teaching in Higher Education*, 1–16. https://doi.org/10.1080/13562517.2021.1928061

Malecka, B., Boud, D., & Carless, D. (2022). Eliciting, processing and enacting feedback: Mechanisms for embedding student feedback literacy within the curriculum. *Teaching in Higher Education*, *27*(7), 908–922. https://doi.org/10.1080/13562517.2020.1754784

Man, D., Chau, M. H., & Kong, B. (2021). Promoting student engagement with teacher feedback through rebuttal writing. *Educational Psychology*, *41*(7), 883–901. https://doi.org/10.1080/01443410.2020.1746238

Manohar, S. G., Chong, T. T.-J., Apps, M. A. J., Batla, A., Stamelou, M., Jarman, P. R., Bhatia, K. P., & Husain, M. (2015). Reward pays the cost of noise reduction in motor and cognitive control. *Current Biology*, *25*(13), 1707–1716. https://doi.org/10.1016/j.cub.2015.05.038

Manwaring, K. C., Larsen, R., Graham, C. R., Henrie, C. R., & Halverson, L. R. (2017). Investigating student engagement in blended learning settings using experience sampling and structural equation modeling. *The Internet and Higher Education*, *35*, 21–33. https://doi.org/10.1016/j.iheduc.2017.06.002

Marsh, H. W., Pekrun, R., Lichtenfeld, S., Guo, J., Arens, A. K., & Murayama, K. (2016). Breaking the double-edged sword of effort/trying hard: Developmental equilibrium and longitudinal relations among effort, achievement, and academic self-concept. *Developmental Psychology*, *52*(8), 1273–1290. https://doi.org/10.1037/dev0000146

Marsh, H. W., Xu, M., & Martin, A. J. (2012). Self-concept: A synergy of theory, method, and application. In K. R. Harris, S. Graham, T. Urdan, C. B. McCormick, G. M. Sinatra, & J. Sweller (Eds.), *APA educational psychology handbook, Vol 1: Theories, constructs, and critical issues* (pp. 427–458). American Psychological Association. https://doi.org/10.1037/13273-015

Martin, F., Chen, Y., Moore, R. L., & Westine, C. D. (2020). Systematic review of adaptive learning research designs, context, strategies, and technologies from 2009 to 2018.

Educational Technology Research and Development, *68*(4), 1903–1929. https://doi.org/10.1007/s11423-020-09793-2

Mason, L. H., Davison, M. D., Hammer, C. S., Miller, C. A., & Glutting, J. J. (2013). Knowledge, writing, and language outcomes for a reading comprehension and writing intervention. *Reading and Writing*, *26*(7), 1133–1158. https://doi.org/10.1007/s11145-012-9409-0

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist*, *32*(1), 1–19. https://doi.org/10.1207/s15326985ep3201_1

Mayer, R. E. (2014). *The Cambridge handbook of multimedia learning* (2nd ed.). Cambridge University Press.

Mayordomo, R. M., Espasa, A., Guasch, T., & Martínez-Melo, M. (2022). Perception of online feedback and its impact on cognitive and emotional engagement with feedback. *Education and Information Technologies*, *27*, 7947–7971. https://doi.org/10.1007/s10639-022-10948-2

McAnulty, S. J. (1981). Paraphrase, summary, precis: Advantages, definitions, models. *Teaching English in the Two-Year College*, *8*(1), 4751.

McCarthy, K. S., & McNamara, D. S. (2021). The multidimensional knowledge in text comprehension framework. *Educational Psychologist*, *56*(3), 196–214. https://doi.org/10.1080/00461520.2021.1872379

Meece, J. L., Wigfield, A., & Eccles, J. S. (1990). Predictors of math anxiety and its influence on young adolescents' course enrollment intentions and performance in mathematics. *Journal of Educational Psychology*, *82*(1), 60–70. https://doi.org/10.1037/0022-0663.82.1.60

Meghji, A. F., Mahoto, N. A., Unar, M. A., & Shaikh, M. A. (2020). The role of knowledge management and data mining in improving educational practices and the learning infrastructure. *Mehran University Research Journal of Engineering and Technology*, *39*(2), 310–323. https://doi.org/10.22581/muet1982.2002.08

Meyer, J., Fleckenstein, J., & Köller, O. (2019). Expectancy value interactions and academic achievement: Differential relationships with achievement measures. *Contemporary Educational Psychology*, *58*, 58–74. https://doi.org/10.1016/j.cedpsych.2019.01.006

Miller, B. W. (2015). Using reading times and eye-movements to measure cognitive engagement. *Educational Psychologist*, *50*(1), 31–42. https://doi.org/10.1080/00461520.2015.1004068

Miller, R. B., DeBacker, T. K., & Greene, B. A. (1999). Perceived instrumentality and academics: The link to task valuing. *Journal of Instructional Psychology*, *26*(4), 250–260.

Mok, W. S. Y., & Chan, W. W. L. (2016). How do tests and summary writing tasks enhance long-term retention of students with different levels of test anxiety? *Instructional Science*, *44*(6), 567–581. https://doi.org/10.1007/s11251-016-9393-x

Molloy, E., Boud, D., & Henderson, M. (2020). Developing a learning-centered framework for feedback literacy. *Assessment & Evaluation in Higher Education*, *45*(4), 527–540. https://doi.org/10.1080/02602938.2019.1667955

Moos, D. C., & Azevedo, R. (2008). Monitoring, planning, and self-efficacy during learning with hypermedia: The impact of conceptual scaffolds. *Computers in Human Behavior*, *24*(4), 1686–1706. https://doi.org/10.1016/j.chb.2007.07.001

Müller, N. M., & Seufert, T. (2018). Effects of self-regulation prompts in hypermedia learning on learning performance and self-efficacy. *Learning and Instruction*, *58*, 1–11. https://doi.org/10.1016/j.learninstruc.2018.04.011

Nagengast, B., Marsh, H. W., Scalas, L. F., Xu, M. K., Hau, K.-T., & Trautwein, U. (2011). Who took the “×” out of expectancy-value theory?: A psychological mystery, a substantive-methodological synergy, and a cross-national generalization. *Psychological Science*, *22*(8), 1058–1066. https://doi.org/10.1177/0956797611415540

Naismith, L. M., Cheung, J. J. H., Ringsted, C., & Cavalcanti, R. B. (2015). Limitations of subjective cognitive load measures in simulation-based procedural training. *Medical Education*, *49*(8), 805–814. https://doi.org/10.1111/medu.12732

Narciss, S. (2006). *Informatives tutorielles Feedback: Entwicklungs- und Evaluationsprinzipien auf der Basis instruktionspsychologischer Erkenntnisse* (Vol. 56). Waxmann.

Narciss, S. (2013). Designing and Evaluating Tutoring Feedback Strategies for digital learning environments on the basis of the Interactive Tutoring Feedback Model. *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. M. Bridges, C. K. Y. Chan, & R. Glofcheski (Eds.), *Scaling up Assessment for Learning in Higher Education* (Vol. 5, pp. 173–189). Springer Singapore.

Narciss, S., Sosnovsky, S., Schnaubert, L., Andrès, E., Eichelmann, A., Goguadze, 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

Nease, A. A., Mudgett, B. O., & Quinones, M. A. (1999). Relationships among feedback sign, self-efficacy, and acceptance of performance feedback. *Journal of Applied Psychology*, *84*(5), 806–814. https://doi.org/10.1037/0021-9010.84.5.806

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

Nieminen, J. H., & Carless, D. (2023). Feedback literacy: A critical review of an emerging concept. *Higher Education*, *85*(6), 1381–1400. https://doi.org/10.1007/s10734-022-00895-9

Nückles, M., Hübner, S., Dümer, S., & Renkl, A. (2010). Expertise reversal effects in writing-to-learn. *Instructional Science*, *38*(3), 237–258. https://doi.org/10.1007/s11251-009-9106-9

Nückles, M., Hübner, S., & Renkl, A. (2009). Enhancing self-regulated learning by writing learning protocols. *Learning and Instruction*, *19*(3), 259–271. https://doi.org/10.1016/j.learninstruc.2008.05.002

Nückles, M., Roelle, J., Glogger-Frey, I., Waldeyer, J., & Renkl, A. (2020). The Self-Regulation-View in Writing-to-Learn: Using Journal Writing to Optimize Cognitive Load in Self-Regulated Learning. *Educational Psychology Review*, *32*(4), 1089–1126. https://doi.org/10.1007/s10648-020-09541-1

OpenAI. (2022). *Introducing ChatGPT*. https://openai.com/index/chatgpt/

Oxford, R. L. (2011). *Teaching and researching language learning strategies: Self-regulation in context* (2nd ed.). Routledge, Taylor & Francis Group.

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

Paas, F. G. W. C., Tuovinen, J. E., Tabbers, H., & van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*, *38*(1), 63–71. https://doi.org/10.1207/S15326985EP3801_8

Paas, F. G. W. C., Tuovinen, J. E., van Merriënboer, J. J. G., & Darabi, A. A. (2005). A motivational perspective on the relation between mental effort and performance: Optimizing learner involvement in instruction. *Educational Technology Research and Development*, *53*(3), 25–34. https://doi.org/10.1007/BF02504795

Paas, F. G. W. C., van Merriënboer, J. J., & Adam, J. J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills*, *79*(1), 419–430. https://doi.org/10.2466/pms.1994.79.1.419

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

Panadero, E., Jonsson, A., & Botella, J. (2017). Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses. *Educational Research Review*, *22*, 74–98. https://doi.org/10.1016/j.edurev.2017.08.004

Pearson, W. S. (2024). Affective, behavioural, and cognitive engagement with written feedback on second language writing: A systematic methodological review. *Frontiers in Education*, *9*, 1285954. https://doi.org/10.3389/feduc.2024.1285954

Peers, I. (2006). *Statistical analysis for education and psychology researchers: Tools for researchers in education and psychology*. Taylor and Francis.

Perez, T., Cromley, J. G., & Kaplan, A. (2014). The role of identity development, values, and costs in college STEM retention. *Journal of Educational Psychology*, *106*(1), 315–329. https://doi.org/10.1037/a0034027

Perez, T., Dai, T., Kaplan, A., Cromley, J. G., Brooks, W. D., White, A. C., Mara, K. R., & Balsai, M. J. (2019). Interrelations among expectancies, task values, and perceived costs in undergraduate biology achievement. *Learning and Individual Differences*, *72*, 26–38. https://doi.org/10.1016/j.lindif.2019.04.001

Perez, T., Gregory, K. H., & Baker, P. B. (2022). Brief research report: Pilot testing an identity-based relevance-writing intervention to support developmental community college students' persistence. *The Journal of Experimental Education*, *90*(1), 77–87. https://doi.org/10.1080/00220973.2020.1800562

Perin, D. (2019). Reading, writing, and self-efficacy of low-skilled postsecondary students. In D. Perin (Ed.), *The Wiley handbook of adult literacy* (1st ed., pp. 237–260). Wiley. https://doi.org/10.1002/9781119261407.ch11

Perin, D., Lauterbach, M., Raufman, J., & Kalamkarian, H. S. (2017). Text-based writing of low-skilled postsecondary students: Relation to comprehension, self-efficacy, and teacher judgments. *Reading and Writing*, *30*(4), 887–915. https://doi.org/10.1007/s11145-016-9706-0

Piaget, J. (1950). *The psychology of intelligence*. Harcourt Brace Jovanovich.

Plass, J. L., & Kalyuga, S. (2019). Four ways of considering emotion in cognitive load theory. *Educational Psychology Review*, *31*(2), 339–359. https://doi.org/10.1007/s10648-019-09473-5

Plass, J. L., Moreno, R., & Brünken, R. (2010). *Cognitive load theory*. Cambridge University Press.

Price, M., Handley, K., & Millar, J. (2011). Feedback: Focusing attention on engagement. *Studies in Higher Education*, *36*(8), Article 8. https://doi.org/10.1080/03075079.2010.483513

Proske, A., Narciss, S., & McNamara, D. S. (2012). Computer-based scaffolding to facilitate students' development of expertise in academic writing. *Journal of Research in Reading*, *35*(2), 136–152. https://doi.org/10.1111/j.1467-9817.2010.01450.x

Putwain, D. W., Nicholson, L. J., Pekrun, R., Becker, S., & Symes, W. (2019). Expectancy of success, attainment value, engagement, and achievement: A moderated mediation analysis. *Learning and Instruction*, *60*, 117–125. https://doi.org/10.1016/j.learninstruc.2018.11.005

Rinehart, S. D., Stahl, S. A., & Erickson, L. G. (1986). Some effects of summarization training on reading and studying. *Reading Research Quarterly*, *21*(4), 422. https://doi.org/10.2307/747614

Rivard, L. P. (2001). Summary writing: A multi-grade study of French-immersion and francophone secondary students. *Language, Culture and Curriculum*, *14*(2), 171–186. https://doi.org/10.1080/07908310108666620

Roelle, J., Müller, C., Roelle, D., & Berthold, K. (2015). Learning from instructional explanations: Effects of prompts based on the active-constructive-interactive framework. *PLOS ONE*, *10*(4), e0124115. https://doi.org/10.1371/journal.pone.0124115

Roelle, J., Nowitzki, C., & Berthold, K. (2017). Do cognitive and metacognitive processes set the stage for each other? *Learning and Instruction*, *50*, 54–64. https://doi.org/10.1016/j.learninstruc.2016.11.009

Roscoe, R. D., & Chi, M. T. H. (2008). Tutor learning: The role of explaining and responding to questions. *Instructional Science*, *36*(4), 321–350. https://doi.org/10.1007/s11251-007-9034-5

Roscoe, R. D., Snow, E. L., Allen, L. K., & McNamara, D. S. (2015). Automated detection of essay revising patterns: Applications for intelligent feedback in a writing tutor. *Technology, Instruction, Cognition and Learning*, *10*, 59–79.

Rosenzweig, E. Q., Wigfield, A., & Eccles, J. S. (2019). Expectancy-value theory and its relevance for student motivation and learning. In K. A. Renninger & S. E. Hidi (Eds.), *The Cambridge handbook of motivation and learning* (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781316823279

Rosenzweig, E. Q., Wigfield, A., & Eccles, J. S. (2022). Beyond utility value interventions: The why, when, and how for next steps in expectancy-value intervention research. *Educational Psychologist*, *57*(1), 11–30. https://doi.org/10.1080/00461520.2021.1984242

Rosenzweig, E. Q., Wigfield, A., & Hulleman, C. S. (2020). More useful or not so bad? Examining the effects of utility value and cost reduction interventions in college physics. *Journal of Educational Psychology*, *112*(1), 166–182. https://doi.org/10.1037/edu0000370

Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology*, *61*, 101860. https://doi.org/10.1016/j.cedpsych.2020.101860

Sadler, D. R. (2006). Formative Assessment: Revisiting the territory. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 77–84. https://doi.org/10.1080/0969595980050104

Safavian, N. (2019). What makes them persist? Expectancy-value beliefs and the math participation, performance, and preparedness of Hispanic youth. *AERA Open*, *5*(3), 233285841986934. https://doi.org/10.1177/2332858419869342

Safavian, N., & Conley, A. (2016). Expectancy-value beliefs of early-adolescent Hispanic and Non-Hispanic youth: Predictors of mathematics achievement and enrollment. *AERA Open*, *2*(4), 233285841667335. https://doi.org/10.1177/2332858416673357

Salomon, G. (1983). The differential investment of mental effort in learning from different sources. *Educational Psychologist*, *18*(1), 42–50. https://doi.org/10.1080/00461528309529260

Schmeck, A., Opfermann, M., van Gog, T., Paas, F. G. W. C., & Leutner, D. (2015). Measuring cognitive load with subjective rating scales during problem solving: Differences between immediate and delayed ratings. *Instructional Science*, *43*(1), Article 1. https://doi.org/10.1007/s11251-014-9328-3

Schmid, R., Pauli, C., Stebler, R., Reusser, K., & Petko, D. (2022). Implementation of technology-supported personalized learning—Its impact on instructional quality. *The*

Journal of Educational Research, 1–12. https://doi.org/10.1080/00220671.2022.2089086

Schnotz, W. (2006). Was geschieht im Kopf des Lesers? Mentale Konstruktionsprozesse beim Textverstehen aus der Sicht der Psychologie und der kognitiven Linguistik. *Text-Verstehen. Grammatik Und Darüber Hinaus*, 222–238.

Schoonen, R. (2019). Are reading and writing building on the same skills? The relationship between reading and writing in L1 and EFL. *Reading and Writing*, *32*(3), 511–535. https://doi.org/10.1007/s11145-018-9874-1

Schunk, D. H., Pintrich, P. R., & Meece, J. L. (2008). *Motivation in education* (3rd ed.). Pearson.

Schunk, D. H., & Rice, J. M. (1991). Learning goals and progress feedback during reading comprehension instruction. *Journal of Reading Behavior*, *23*(3), 351–364. https://doi.org/10.1080/10862969109547746

Seifried, E., Lenhard, W., & Spinath, B. (2015). Plagiarism detection: A comparison of teaching assistants and a software tool in identifying cheating in a psychology course. *Psychology Learning & Teaching*, *14*(3), 236–249. https://doi.org/10.1177/1475725715617114

Seifried, E., Lenhard, W., & Spinath, B. (2016). Automatic essay assessment: Effects on students' acceptance and on learning-related characteristics. *Psihologija*, *49*(4), 469–482. https://doi.org/10.2298/PSI1604469S

Seifried, E., Lenhard, W., & Spinath, B. (2017). Filtering essays by means of a software tool: Identifying poor essays. *Journal of Educational Computing Research*, *55*(1), 26–45. https://doi.org/10.1177/0735633116652407

Shanahan, T. (2019). Reading—Writing connections. In S. Graham, C. A. McArthur, & M. Hebert (Eds.), *Best practices in writing instruction*. Guilford Press.

Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., & Botvinick, M. M. (2017). Toward a rational and mechanistic account of mental effort. *Annual Review of Neuroscience*, *40*(1), 99–124. https://doi.org/10.1146/annurev-neuro-072116-031526

Shi, H., & Aryadoust, V. (2024). A systematic review of AI-based automated written feedback research. *ReCALL*, *36*(2), 187–209. https://doi.org/10.1017/S0958344023000265

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153–189. https://doi.org/10.3102/0034654307313795

Silver, W. S., Mitchell, T. R., & Gist, M. E. (1995). Responses to successful and unsuccessful performance: The moderating effect of self-efficacy on the relationship between performance and attributions. *Organizational Behavior and Human Decision Processes*, *62*(3), 286–299. https://doi.org/10.1006/obhd.1995.1051

Stevens, E. A., Park, S., & Vaughn, S. (2019). A review of summarizing and main idea interventions for struggling readers in grades 3 through 12: 1978–2016. *Remedial and Special Education*, *40*(3), 131–149. https://doi.org/10.1177/0741932517749940

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 & Education*, *131*, 33–48. https://doi.org/10.1016/j.compedu.2018.12.005

Sun, J. C.-Y., & Rueda, R. (2012). Situational interest, computer self-efficacy and self-regulation: Their impact on student engagement in distance education. *British Journal of Educational Technology*, *43*(2), 191–204. https://doi.org/10.1111/j.1467-8535.2010.01157.x

Sun, J. C.-Y., Yu, S.-J., & Chao, C.-H. (2019). Effects of intelligent feedback on online learners' engagement and cognitive load: The case of research ethics education. *Educational Psychology*, *39*(10), Article 10. https://doi.org/10.1080/01443410.2018.1527291

Sung, Y.-T., Liao, C.-N., Chang, T.-H., Chen, C.-L., & Chang, K.-E. (2016). The effect of online summary assessment and feedback system on the summary writing on 6th graders: The LSA-based technique. *Computers & Education*, *95*, 1–18. https://doi.org/10.1016/j.compedu.2015.12.003

Sutton, P. (2012). Conceptualizing feedback literacy: Knowing, being, and acting. *Innovations in Education and Teaching International*, *49*(1), 31–40. https://doi.org/10.1080/14703297.2012.647781

Sweller, J. (2011). Cognitive Load Theory. In Mestre, Jose P. & Ross, Brian H. (Eds.), *Psychology of learning and motivation* (Vol. 55, pp. 37–76). Academic Press. https://doi.org/10.1016/B978-0-12-387691-1.00002-8

Sweller, J. (2020). Cognitive load theory and educational technology. *Educational Technology Research and Development*, *68*(1), 1–16. https://doi.org/10.1007/s11423-019-09701-3

Taboada, A., Tonks, S. M., Wigfield, A., & Guthrie, J. (2009). Effects of motivational and cognitive variables on reading comprehension. *Reading and Writing*, *22*(1), 85. https://doi.org/10.1007/s11145-008-9133-y

Tay, H. Y., & Lam, K. W. L. (2022). Students' engagement across a typology of teacher feedback practices. *Educational Research for Policy and Practice*, *21*(3), 427–445. https://doi.org/10.1007/s10671-022-09315-2

Teng, M. F. (2022). Effects of individual and group metacognitive prompts on tertiary-level students' metacognitive awareness and writing outcomes. *The Asia-Pacific Education Researcher*, *31*(5), 601–612. https://doi.org/10.1007/s40299-021-00611-8

Tian, L., & Zhou, Y. (2020). Learner engagement with automated feedback, peer feedback and teacher feedback in an online EFL writing context. *System*, *91*, 102247. https://doi.org/10.1016/j.system.2020.102247

To, J. (2022). Using learner-centred feedback design to promote students' engagement with feedback. *Higher Education Research & Development*, *41*(4), Article 4. https://doi.org/10.1080/07294360.2021.1882403

Tolosa, C., East, M., & Villers, H. (2015). Motivating twenty-first-century learners: The impact of an online reciprocal peer-tutoring initiative for foreign language learning. In C. Koh (Ed.), *Motivation, Leadership and Curriculum design* (pp. 137–149). Springer Singapore. https://doi.org/10.1007/978-981-287-230-2_11

Trautwein, U., Marsh, H. W., Nagengast, B., Lüdtke, O., Nagy, G., & Jonkmann, K. (2012). Probing for the multiplicative term in modern expectancy–value theory: A latent interaction modeling study. *Journal of Educational Psychology*, *104*(3), 763–777. https://doi.org/10.1037/a0027470

Updegraff, K. A., Eccles, J. S., Barber, B. L., & O'Brien, K. M. (1996). Course enrollment as self-regulatory behavior: Who takes optional high school math courses? *Learning and Individual Differences*, *8*(3), 239–259. https://doi.org/10.1016/S1041-6080(96)90016-3

van den Boom, G., Paas, F. G. W. C., & van Merriënboer, J. J. G. (2007). Effects of elicited reflections combined with tutor or peer feedback on self-regulated learning and learning outcomes. *Learning and Instruction*, *17*(5), 532–548. https://doi.org/10.1016/j.learninstruc.2007.09.003

van den Boom, G., Paas, F. G. W. C., van Merriënboer, J. J. G., & van Gog, T. (2004). Reflection prompts and tutor feedback in a web-based learning environment: Effects on students' self-regulated learning competence. *Computers in Human Behavior*, *20*(4), 551–567. https://doi.org/10.1016/j.chb.2003.10.001

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 der Kleij, F. M., & Lipnevich, A. A. (2021). Student perceptions of assessment feedback: A critical scoping review and call for research. *Educational Assessment, Evaluation and Accountability*, *33*(2), 345–373. https://doi.org/10.1007/s11092-020-09331-x

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. Academic Press.

van Lehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, *46*(4), 197–221. https://doi.org/10.1080/00461520.2011.611369

Vosniadou, S. (Ed.). (2008). *International handbook of research on conceptual change* (1st ed.). Routledge.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Wade-Stein, D., & Kintsch, E. (2004). Summary Street: Interactive computer support for writing. *Cognition and Instruction*, *22*(3), 333–362. https://doi.org/10.1207/s1532690xci2203_3

Wahl, D. (2001). Nachhaltige Wege vom Wissen zum Handeln. *Beiträge zur Lehrerbildung*, *19*(2), 157–174. https://doi.org/10.25656/01:13453

Wang, W. (2014). Students' perceptions of rubric-referenced peer feedback on EFL writing: A longitudinal inquiry. *Assessing Writing*, *19*, 80–96. https://doi.org/10.1016/j.asw.2013.11.008

Wäschle, K., Gebhard, A., Oberbusch, E.-M., & Nückles, M. (2015). Journal writing in science: Effects on comprehension, interest, and critical reflection. *Journal of Writing Research*, *7*(1), 41–64. https://doi.org/10.17239/jowr-2015.07.01.03

Watt, H. M. G., Shapka, J. D., Morris, Z. A., Durik, A. M., Keating, D. P., & Eccles, J. S. (2012). Gendered motivational processes affecting high school mathematics participation, educational aspirations, and career plans: A comparison of samples from Australia, Canada, and the United States. *Developmental Psychology*, *48*(6), 1594–1611. https://doi.org/10.1037/a0027838

Westby, C., Culatta, B., Lawrence, B., & Hall-Kenyon, K. (2010). Summarizing expository texts. *Topics in Language Disorders*, *30*(4), 275–287. https://doi.org/10.1097/TLD.0b013e3181ff5a88

Wigfield, A., & Cambria, J. (2010a). Expectancy-value theory: Retrospective and prospective. In T. C. Urdan & S. A. Karabenick (Eds.), *Advances in Motivation and Achievement* (Vol. 16, pp. 35–70). Emerald Group Publishing Limited. https://doi.org/10.1108/S0749-7423(2010)000016A005

Wigfield, A., & Cambria, J. (2010b). Students' achievement values, goal orientations, and interest: Definitions, development, and relations to achievement outcomes. *Developmental Review*, *30*(1), 1–35. https://doi.org/10.1016/j.dr.2009.12.001

Wigfield, A., & Eccles, J. S. (2000). Expectancy–value theory of achievement motivation. *Contemporary Educational Psychology*, *25*(1), Article 1. https://doi.org/10.1006/ceps.1999.1015

Wigfield, A., & Eccles, J. S. (2002). The development of competence beliefs, expectancies for success, and achievement values from childhood through adolescence. In *Development of Achievement Motivation* (pp. 91–120). Elsevier. https://doi.org/10.1016/B978-012750053-9/50006-1

Wigfield, A., Eccles, J. S., Yoon, K. S., Harold, R. D., Arbreton, A. J. A., Freedman-Doan, C., & Blumenfeld, P. C. (1997). Change in children's competence beliefs and subjective task values across the elementary school years: A 3-year study. *Journal of Educational Psychology*, *89*(3), 451–469. https://doi.org/10.1037/0022-0663.89.3.451

Wilson, J., Olinghouse, N. G., & Andrada, G. N. (2014). Does automated feedback improve writing quality? *Learning Disabilities--A Contemporary Journal*, *12*(1), 93–118.

Winstone, N. E., Mathlin, G., & Nash, R. A. (2019). Building feedback literacy: Students' perceptions of the developing engagement with feedback toolkit. *Frontiers in Education*, *4*, 39. https://doi.org/10.3389/feduc.2019.00039

Winstone, N. E., Nash, R. A., Parker, M., & Rowntree, J. (2017). Supporting learners' agentic engagement with feedback: A systematic review and a taxonomy of recipience processes. *Educational Psychologist*, *52*(1), 17–37. https://doi.org/10.1080/00461520.2016.1207538

Wischgoll, A. (2016). Combined training of one cognitive and one metacognitive strategy improves academic writing skills. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00187

Wischgoll, A. (2017). Improving undergraduates' and postgraduates' academic writing skills with strategy training and feedback. *Frontiers in Education*, *2*, 33. https://doi.org/10.3389/feduc.2017.00033

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

Wu, Y., & Kang, X. (2021). A moderated mediation model of expectancy-value interactions, engagement, and foreign language performance. *SAGE Open*, *11*(4), 215824402110591. https://doi.org/10.1177/21582440211059176

Yang, J., Tan, Q., Tang, Y., & Bai, C. (2021). Effect of embedding prompts on learning performance and metacognitive monitoring. *2021 IEEE International Conference on Engineering, Technology & Education (TALE)*, 01–07. https://doi.org/10.1109/TALE52509.2021.9678745

Yee, D. M., & Braver, T. S. (2018). Interactions of motivation and cognitive control. *Current Opinion in Behavioral Sciences*, *19*, 83–90. https://doi.org/10.1016/j.cobeha.2017.11.009

Yossatorn, Y., Awuor, N. O., & Weng, C. (2024). The relationships among online self-regulated english learning, task value, and academic self-efficacy: The mediating role of the task value. *Current Psychology*, *43*(17), 15705–15724. https://doi.org/10.1007/s12144-023-05524-x

Yu, S., Zhang, Y., Zheng, Y., Yuan, K., & Zhang, L. (2019). Understanding student engagement with peer feedback on master's theses: A Macau study. *Assessment & Evaluation in Higher Education*, *44*(1), 50–65. https://doi.org/10.1080/02602938.2018.1467879

Yu, S., Zhou, N., Zheng, Y., Zhang, L., Cao, H., & Li, X. (2019). Evaluating student motivation and engagement in the Chinese EFL writing context. *Studies in Educational Evaluation*, *62*, 129–141. https://doi.org/10.1016/j.stueduc.2019.06.002

Zellermayer, M., Salomon, G., Globerson, T., & Givon, H. (1991). Enhancing writing-related metacognitions through a computerized writing partner. *American Educational Research Journal*, *28*(2), 373–391.

Zhai, N., & Ma, X. (2021). Automated writing evaluation (AWE) feedback: A systematic investigation of college students' acceptance. *Computer Assisted Language Learning*, 1–26. https://doi.org/10.1080/09588221.2021.1897019

Zhang, S., & Liu, Q. (2019). Investigating the relationships among teachers' motivational beliefs, motivational regulation, and their learning engagement in online professional learning communities. *Computers & Education*, *134*, 145–155. https://doi.org/10.1016/j.compedu.2019.02.013

Zhang, Y., & Gao, Y. (2024). Exploring the dynamics of student engagement with receiving peer feedback in L2 writing. *Assessing Writing*, *60*, 100842. https://doi.org/10.1016/j.asw.2024.100842

Zhang, Z. (2020). Engaging with automated writing evaluation (AWE) feedback on L2 writing: Student perceptions and revisions. *Assessing Writing*, *43*, 100439. https://doi.org/10.1016/j.asw.2019.100439

Zhang, Z., & Hyland, K. (2022). Fostering student engagement with feedback: An integrated approach. *Assessing Writing*, *51*, 100586. https://doi.org/10.1016/j.asw.2021.100586

Zheng, L. (2016). The effectiveness of self-regulated learning scaffolds on academic performance in computer-based learning environments: A meta-analysis. *Asia Pacific Education Review*, *17*(2), 187–202. https://doi.org/10.1007/s12564-016-9426-9

Zheng, Y., & Yu, S. (2018). Student engagement with teacher written corrective feedback in EFL writing: A case study of Chinese lower-proficiency students. *Assessing Writing*, *37*, 13–24. https://doi.org/10.1016/j.asw.2018.03.001

Zimbardi, K., Colthorpe, K., Dekker, A., Engstrom, C., Bugarcic, A., Worthy, P., Victor, R., Chunduri, P., Lluka, L., & Long, P. (2017). Are they using my feedback? The extent of students' feedback use has a large impact on subsequent academic performance. *Assessment & Evaluation in Higher Education*, *42*(4), 625–644. https://doi.org/10.1080/02602938.2016.1174187

Zimmerman, B. J. (2000). Attaining self-regulation. In M. Boekaerts, Pintrich, Paul R., & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). Academic Press. https://doi.org/10.1016/B978-012109890-2/50031-7

Zimmerman, B. J., & Kitsantas, A. (2007). A writer's discipline: The development of self-regulatory skill. In G. Rijlaarsdam, P. Boscolo, & S. Hidi (Eds.), *Studies in Writing* (Vol. 19, pp. 51–69). Elsevier.

Appendix

- Manuscripts
  - Barkela, V., & Leuchter, M. (2024). Effectiveness of automated formative feedback in an online tutorial for promoting summarizing. *Journal of Educational Technology Development and Exchange, 17*(1), 62-90. doi.org/10.18785/jetde.1701.04
  - Barkela, V., Schmitt, L., & Leuchter, M. (2023). The impact of cognitive and motivational resources on engagement with automated formative feedback. *Contemporary Educational Psychology*, 102234. https://doi.org/10.1016/j.cedpsych.2023.102234
  - Barkela, V., & Leuchter, M. (2024). The interplay of instructional prompts and automated formative feedback. *Manuscript submitted for publication.*
- Curriculum vitae
- Declaration of originality

Manuscript 1

Barkela, V., & Leuchter, M. (2024). Effectiveness of automated formative feedback in an online tutorial for promoting summarizing. *Journal of Educational Technology Development and Exchange, 17*(1), 62-90. doi.org/10.18785/jetde.1701.04

Impact factor: 1.0

Effectiveness of Automated Formative Feedback in an Online Tutorial for Promoting Summarizing

Veronika Barkela & Miriam Leuchter

ABSTRACT: *We conducted a study with the aim to investigate the effectiveness of automated formative feedback in improving students' ability to summarize. N = 138 undergraduate students in an education program were asked to summarize six scientific texts, with the experimental group (n=87) receiving automated formative feedback in a computer-based learning environment (FALB).* FALB *provides automated feedback about content coverage, copying words avoidance, redundancy avoidance, relevance, and length. Comparing the experimental group to a control group (n = 51), results implied that summarizing skills can be fostered when interacting with* FALB. *In particular, the automated formative feedback promoted the adherence to the predefined length and the avoidance of copying words while maintaining a high content coverage, fostering cognitive processes essential for constructing a mental model of a text. In addition, students in the experimental group were able to maintain high quality summaries in their final session when not scaffolded. In conclusion,* FALB *supports the alignment of internal standards with external standards and provides an incentive to revise and engage with texts.*

Keywords: formative feedback, automated summary evaluator, summary writing, technology- enhanced learning environments, pre-service teacher students

1. Introduction

University students are anticipated to quickly extract and assimilate relevant information from scientific literature and articulate it with precision and conciseness (Kürschner & Schnotz, 2007; van Dijk & Kintsch, 1983). For this purpose, summarizing has been shown to be an adequate learning strategy (Mok & Chan, 2016; Stevens et al., 2019). Summarizing requires understanding a text, identifying relevant aspects, and reflecting on the topic (Perin et al., 2017; Westby et al., 2010). Furthermore, it entails writing a short, concise version of a text, maintaining the key aspects, and formulating them in one's own words, while avoiding redundant and irrelevant parts (Dunlosky et al., 2013; Kirkland & Saunders, 1991). The quality of a summary is highly associated with a person's mental model of the original text (M. K. Kim & McCarthy, 2021; Schnotz, 2006). Mental models are representations of a text and encompass both explicitly stated information from the text and inferences drawn from the text by connecting related information with prior knowledge (van Dijk & Kintsch, 1983). People with more prior knowledge about the topic of a text are more likely to create a comprehensive mental model and write a good summary (K. Kim et al., 2019).

However, students' representations of how to learn with summaries are often incomplete and, they lack effective skills for creating a mental model of a text and summarizing it (Friend, 2001). They tend to simply copy phrases, forgo reflecting on the content of a text, and refrain from condensing the text to its key aspects, thereby depriving themselves of learning effectively from text (Ahn, 2022; Duke & Pearson, 2009). Such behavior implies that students' internal reference standards of a good summary and how to summarize differ from the external standard of a high-quality summary and successful summarizing strategies. Thus, effective summarizing skills do not develop naturally, but must be learned (Ahn, 2022; Keck, 2006). Yet, summarizing strategies are often taught merely in elementary school and are not emphasized in later grades, limiting students' ability to use summarizing as an effective learning strategy (McNamara et al., 2019). Therefore, it is an important endeavor at the university to teach students effective summarizing strategies to help them succeed in their studies. However, providing learning opportunities that allow students to develop internal reference standards according to an external standard and thus improve their summarizing skills is hardly feasible for large classes with limited resources (Allen et al., 2016). Automated feedback systems for summarizing overcome this dilemma by providing the opportunity to assess many students immediately and as often as desired while meeting premises of effective feedback (Deeva et al., 2021; Strobl et al., 2019).

The field of technology-enhanced learning is undergoing rapid transformation, with the integration of generative AI, such as *ChatGPT* (OpenAI et al., 2023), reshaping teaching methods and assessment practices. Nevertheless, a constrained expert system utilizing established natural language processing techniques, like latent semantic analysis, may offer distinct advantages in fostering effective summarizing skills. The development of such a system is less resource-intensive compared to approaches relying on large language models, making it feasible even with limited resources. Furthermore, it enables the creation of a focused system that incorporates an external standard for evaluating students' work, allowing students to align their internal standards accordingly. Such systems have been implemented successfully for the English language (M. K. Kim & McCarthy, 2021; H. Li et al., 2018), French (Lemaire & Dessus, 2001), and Chinese (Sung et al., 2016), demonstrating their effectiveness.

Yet, to the best of our knowledge, such an automated feedback system for promoting university students' summarizing skills in German is still missing. This study seeks to fill this gap by expanding the evidence on the effectiveness of these systems to new languages, samples, and designs. Specifically, our aim is to assess the supportive potential of a German feedback system designed for undergraduate elementary education students. We intend to scrutinize the key aspects of summarizing skills promoted by our tool and explore the potential association between the formative aspect of feedback and the quality of summaries. Our approach involves renewing a German feedback system used in elementary and middle schools, initially focused on reading comprehension (conText; Lenhard et al., 2013). Through this redesign, undergraduate elementary education students can engage in a computer-based learning environment, fostering their summarizing skills and experiencing a tool that they may later employ as elementary school teachers to enhance reading comprehension in their students.

2. Theoretical Background

Summarizing requires multiple cognitive processes, including integrating new information into one's cognitive schema, determining the relevance of information, constructing a mental model of the text, translating the mental model into one's own words, and ultimately writing it down (Hidi & Anderson, 1986; Perin et al., 2017; Westby et al., 2010). Improved coordination of these processes contributes to individuals' proficiency in constructing comprehensive mental models and generating effective summaries (K. Kim et al., 2019).

Certain task designs support the acquisition of effective summarizing skills to better coordinate cognitive processes. For example, not seeing the text at the same time as writing the summary supports information retrieval and prevents word copying and redundancy (Hidi & Anderson, 1986). Moreover, limiting a summary's length encourages condensing content to key messages and deleting irrelevant information (Hill, 1991). Furthermore, various studies have emphasized the supportive role of formative feedback in encouraging iterative revisions, deep text processing, and adherence to task criteria according to an external standard (Graham, 2018; Kellogg & Raulerson, 2007). Several factors contribute to the effectiveness of feedback (Narciss, 2017; Nixon et al., 2016). For example, students derive greater benefits from immediate rather than delayed feedback (Shute, 2008), exhibit enhanced learning outcomes with elaborate as opposed to simple feedback (Hattie & Timperley, 2007), and better align their internal to an external reference standard when receiving individualized versus general feedback (Zhu et al., 2020). Moreover, effective feedback builds on pre-established assessment criteria and previous performance (Black & Wiliam, 2009). The same applies to *automated* formative feedback, which has been shown to be as effective and valid as human feedback (Seifried et al., 2012; Stevenson & Phakiti, 2014; van der Kleij et al., 2015).

Research in both offline and online learning environments has implied that the frequency of revising a draft has major impact on its text quality (J. A. Butler & Britt, 2010; Kirkland & Saunders, 1991; Roscoe et al., 2015; Sung et al., 2016). However, inexperienced writers tend to revise scarcely and superficially (Abba et al., 2018). Providing students with automated formative feedback might strengthen students' engagement in the learning process and encourage more revisions, thus supporting the alignment of internal and external reference standards (Link et al., 2020; Liu et al., 2017). Automated formative feedback can be implemented in a way that allows students to control the amount of feedback they receive by letting the algorithm evaluate their drafts as many times as they want. The number of feedback loops can be an indicator of the intensity of revision, thus positively affecting text quality.

The distinction between automated summary evaluators and automated writing evaluation has not always been clear and technological advancements have overlapped. However, for the purposes of our study, we will mainly focus on the development of advancements in the realm of summarizing. Over the years, several automated summary evaluators have been developed. One notable system, *Summary Street* by Wade-Stein & Kintsch (2004) was among the first to give feedback on summaries to elementary and middle school students, originally designed to enhance text comprehension. They followed a latent semantic analysis approach, which is a natural language processing technique to represent the content of texts. Their English-based computer-based learning environment included source texts about science topics, a text editor for summary composition, bar chart feedback on summary length and section coverage, and a redundancy and relevance check, that listed problematic sentences. The effectiveness of *Summary Street* was investigated using a with-in subject design with counterbalanced order of conditions. One condition provided feedback on length and spelling, while the other supplemented feedback on content coverage, redundancy, and relevance. The results indicated that automated feedback on content coverage significantly assisted students in enhancing the substance of their summaries. Based on this work, Lenhard et al. (2012) developed a similar system for German elementary school students. Their investigation into the effects of this automated summary evaluator revealed positive impacts on students' reading comprehension and fluency compared to control groups.

Sung et al. (2016) conducted a study with Chinese elementary school students to compare the supportive potential of a summary evaluator providing semantic feedback based on text similarity in one condition (Foltz et al., 1999) and concept feedback based on concept maps in another condition (Schvaneveldt & Cohen, 2010). Results suggested positive effects of both semantic feedback and concept feedback on content coverage of the summaries. Furthermore, a decreasing submission count on the posttest indicated that students learned summarizing skills and did not rely on the support tool. Chew et al. (2019) developed an automated summary evaluator for undergraduate computer science students to learn and practice summarizing in the context of foreign language learning. They included concept maps, worked examples, and feedback on summarizing strategies, demonstrating positive effects on the improvement of the summaries' text quality (rated by teachers) from pretest to posttest. Despite these advancements, we have identified research gaps, specifically in our pursuit of promoting effective summarizing skills to German undergraduate elementary education students.

While existing systems have been predominantly used in school settings (Lenhard et al., 2012; Sung et al., 2016), or, if in a university setting, for foreign language learning (Chew et al., 2019), an environment dedicated to explicit practice in processing scientific texts and communicating the information precisely and concisely remains undeveloped. Hence, this study introduces a computer-based learning environment designed to present short German scientific texts related to pedagogical content knowledge to elementary education students. The system offers automated feedback on content coverage and writing style aiming to provide learning opportunities for the development of effective summarizing strategies.

Various automated summary evaluators provide users with information on semantic similarity measures or concept maps of text content (M. K. Kim & McCarthy, 2021; Lenhard et al., 2012; Sung et al., 2016). In contrast, our study focuses on developing metrics aligned with the cognitive processes inherent in summarizing, including the identification of relevant information and the creation of a condensed and concise version of the text in one's own words. Therefore, in addition to details about content coverage and length, the automated feedback encompasses information on the avoidance of copied words, redundancy, and irrelevance.

Automated summary evaluators, designed to provide formative feedback, are intended to encourage students to consistently revise their drafts. Sung et al. (2016) utilized the quantity of feedback loops as a metric for tool utilization, indicative student engagement. Yet, to the best of our knowledge, the correlation between an increased number of feedback loops and the generation of higher-quality summaries remains unexplored. Therefore, our study seeks to elucidate the relationship between the number of feedback loops and the quality of summaries.

Methodologically, while many studies assess the effectiveness of automated formative feedback through posttest-to-pretest comparisons (Chew et al., 2019) or via case studies (M. K. Kim & McCarthy, 2021; Zhang, 2020), our approach involves evaluating change over multiple time points, employing learning trajectories. These trajectories depict students' probable cognitive developments as they progress in their task (Sztajn et al., 2012). This approach enables teachers to make diagnostic inferences and offer tailored feedback based on the data supplied, even with large sample sizes (Beese, 2019; Plass & Pawar, 2020). In computer-based contexts, understanding learning trajectories can help to anticipate learner behavior at different learning stages, designing customized learning environment, and implementing measures for additional support, such as individualized feedback (Lee & Tan, 2017; Schmid et al., 2022). In the following, we will describe the computer-based learning environment *FALB* and outline our pedagogical considerations.

3. FALB

The computer-based learning environment *FALB* was developed to provide learning opportunities for developing more sophisticated summarizing skills. It is based on principles of formative assessment (Black & Wiliam, 2009, 2018). *FALB* is composed of two main components (front end and server, Figure 1) which will be explained in the following.

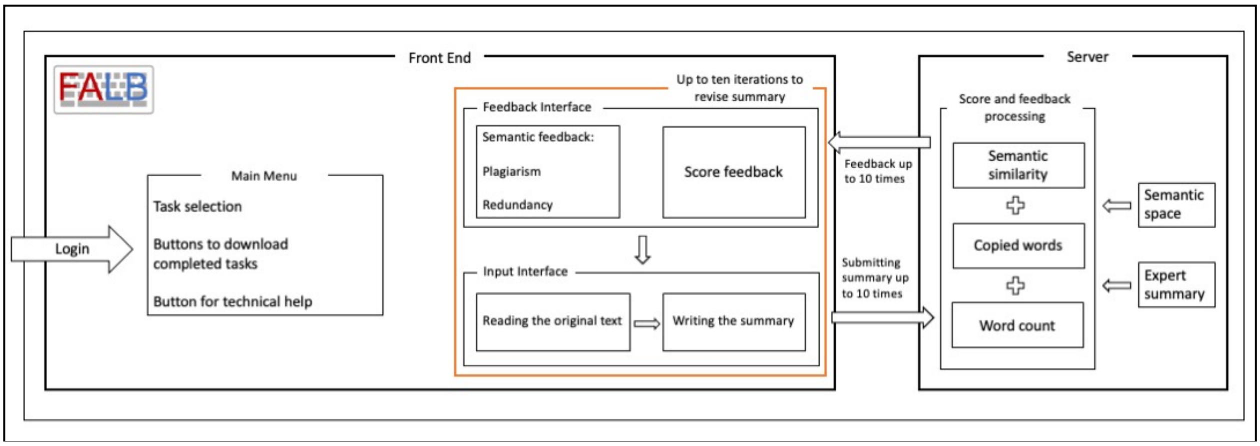


Figure 1
Framework of the Computer-based Learning Environment FALB

3.1. Front End

The front end describes the platform which the user sees and interacts with. A two-page input interface presents the text to be summarized and a text box where the summary can be written down. The original text is structured in two text sections and intentionally displayed separately from the text box to encourage the creation of a mental model and one's own phrasing (Hidi & Anderson, 1986). Re-reading the original text is not limited, the text can be displayed by pressing a button. The feedback Interface displays the feedback and was taken from *conText* (Lenhard et al., 2013) which is based on 'summary street' (Wade-Stein & Kintsch, 2004). Automated formative feedback can be solicited up to ten times per text and session and entails semantic feedback and score feedback.

Semantic feedback provides information about copied words from the original text. Passages are marked as copied and underlined in red when three or more consecutive words are copied from the original text. Additionally, the system provides information about repeated expressions of the same idea (redundancy underlined in blue; Figure 2) and specifies this information in more detail in a pop-up window, listing corresponding sentences with similar information in the same color. Furthermore, it provides information about irrelevant sentences which should help students stay with the text content (irrelevance underlined in grey). Moreover, unknown words are underlined because they are an indicator of spelling mistakes. Semantic feedback can be obtained by pressing a button labeled "submit text".

Score feedback provides information on how well the original text is covered for the three text sections separately. It also provides information on how well copied words or passages are avoided and how well repeating information is avoided. Moreover, it provides information about the length of the summary, which should not exceed 30% of the original text to obtain the maximum score. All scores are displayed in percentages as horizontal bars. They can be obtained by pressing a button and up to ten times if desired (Figure 3). In the following, those feedback scores will be referred to as text quality scores.

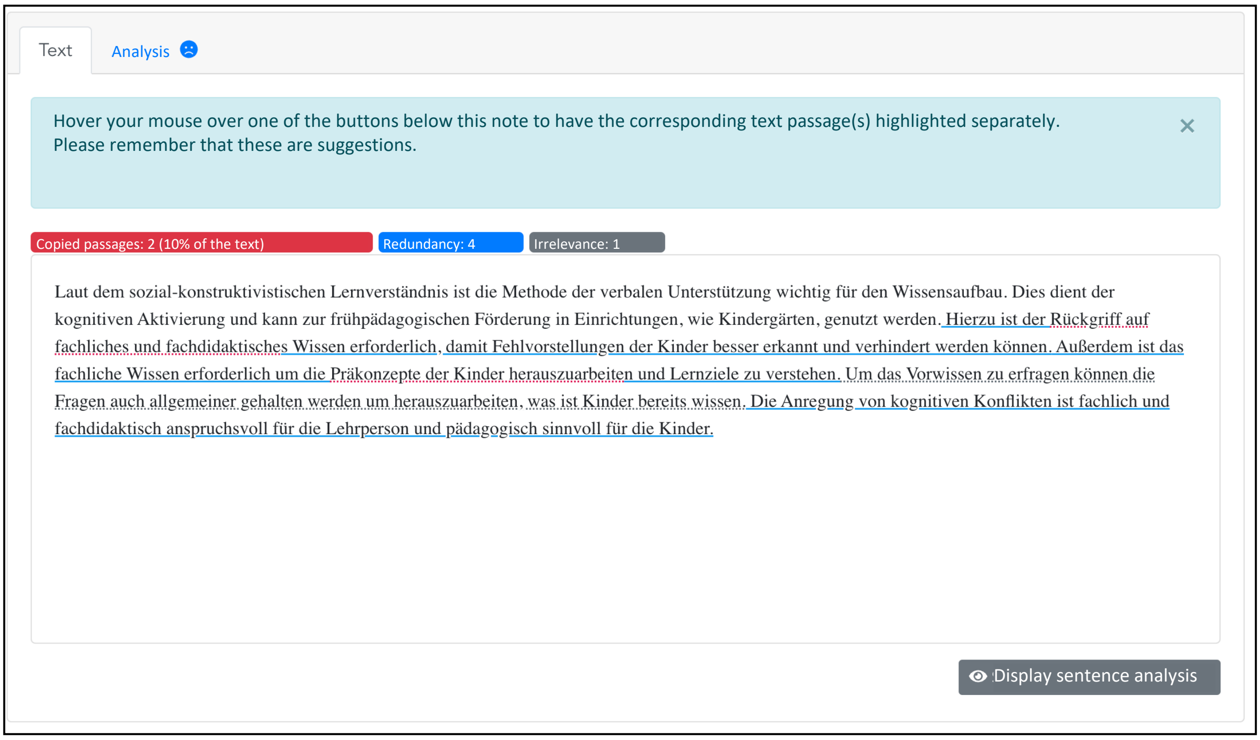


Figure 2

Example of Semantic Feedback

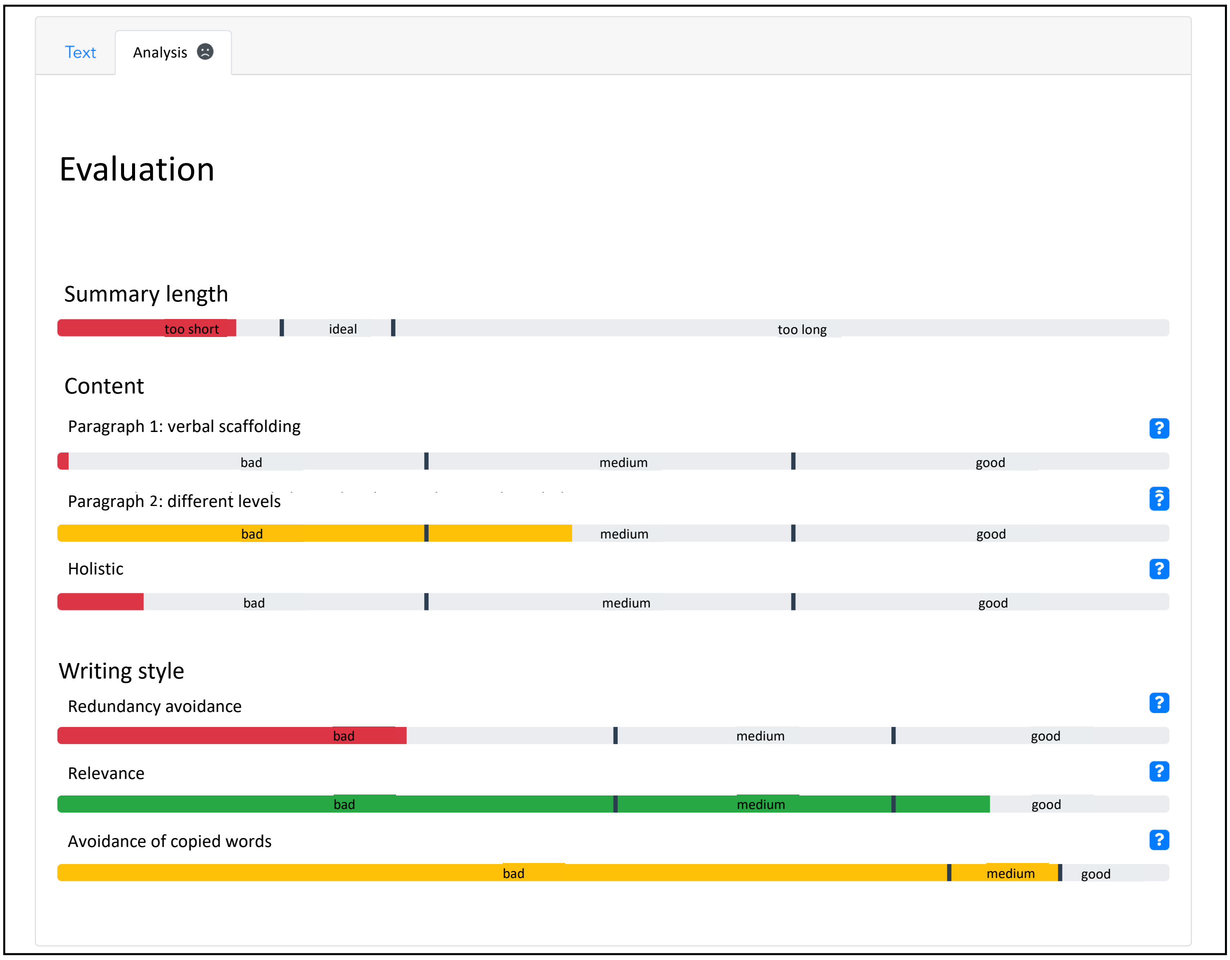


Figure 3
Example of Score Feedback

3.2. Server

The server's main task is to evaluate the summaries and to provide semantic feedback and text quality scores. For this purpose, the original texts, expert summaries, and semantic space were implemented on the server for calculation with Latent Semantic Analysis.

Latent Semantic Analysis (LSA): Text quality scores were determined by LSA, which identifies and sorts words based on their context (Deerwester et al., 1990; Foltz et al., 1999; Landauer, T. K. et al., 1998; Lenhard et al., 2007). LSA requires a large text corpus (semantic space). LSA projects sentences from the original text, the expert summary, and the student summary into the semantic space and computes a vector for each sentence. Based on the similarity of the vectors, LSA can determine the similarity between pairs of sentences. We used the LSA syntax of conText (Lenhard et al., 2013) and fed the corpus with 201,288 different meaningful German words about learning support in science education through teacher-student interaction.

Text Material: Six excerpts from German scientific texts of comparable length (445-649 words) were used and provided in a fixed order. All texts informed about teacher-student interaction in science education. The texts were selected to be at a similar level of readability, as determined by experts and a readability index (LIX: 66.0 – 80.2; Lenhard & Lenhard, 2014).

Expert Summaries: Two experts prepared summaries of the six texts used in this research. Both experts were specialists in teacher education and academic writing. After they each wrote a first draft, they discussed, revised, and combined the drafts according to the summary criteria (content, avoidance of copied words, avoidance of redundancy, relevance, and length).

Calculation of the Text Quality Scores: Content was calculated by comparing the relation of students' summary content coverage to the original text with the relation of the expert's summary content coverage to the original text. Avoidance of copied words was defined as the ratio of non-copied sentences or phrases to copied sentences or phrases. Redundancy avoidance was calculated based on the number of sentences containing repeated information. Relevance was defined as the ratio of irrelevant sentences to sentences containing relevant information. Length was measured as the ratio of the number of words in the summary to the number of words in the original text. All five text quality scores can theoretically range between 0 and 100%. While scores closer to 100% are desirable for content, avoidance of copied words, redundancy avoidance, and relevance, the optimal score for length is between 20 and 30%.

Summary Result: The text quality scores are interdependent. For example, a summary is more likely to capture the entire content of the original text if it is relatively long. However, the longer the summary, the more likely it is to contain irrelevant and redundant parts. Avoiding copied words or passages is easier if the content can be paraphrased, but at the expense of avoiding redundancy. If more redundancy is avoided, the content must be highly condensed, which may result in less content coverage. Therefore, to make the summaries comparable, we calculated an overall summary result that considers all these text quality scores in terms of their importance for the processing of a text. Garner (1982), Head et al. (1989)), and Sung et al. (2016) proposed formulas to quantify text quality. We modified those formulas to include our text quality scores and assigned different weights depending on the importance we ascribed to them.

Content is weighted with the factor 0.5, avoidance of copied words is weighted with the factor 0.3, redundancy avoidance is weighted with the factor 0.15, and relevance is weighted with the factor 0.05. All these factors strongly contribute to the creation of mental models that indicate the transformation of the information in the text into individual knowledge (Schnotz, 2006; van Dijk & Kintsch, 1983). Following Lenhard et al. (2013), the optimal length of a summary was set between 20 and 30% of the original text; thus, the length in the formula is the ratio of the length of the students' summaries and the set length limit.

Content, avoidance of copied words, redundancy avoidance, and relevance add up in the formula's counter to display the text quality as a sum value. The high weighting of content derives from the importance of including all essential aspects of the original text in a summary that demonstrates a thorough understanding of the topic. Avoidance of copied words is also highly weighted and shows that the students can express thoughts in their own words. Redundancy avoidance contributes to a brief and concise presentation of the content, which is an essential aspect of a summary. This score is weighted less because it is relative to the length in the denominator. Relevance is weighted lightly because aspects of relevance are covered in the content factor. Additionally, most of the participants scored very high on the relevance score (85.1% > 90), indicating that students generally have no difficulty including relevant information. The longer the summary, the more likely it is that content and avoidance of copied words will score high and redundancy avoidance and relevance will score low. Therefore, the sum of the content score, the copied word avoidance score, the redundancy avoidance score, and the relevance score is divided by the ratio of the student's summary length to the length limit (cf. Sung et al., 2016). Our formula is as follows:

$$\text{Summary result} = \boxed{\left(\frac{0.5*CT+0.3*CWA+0.15*RA+0.05*RV}{\frac{SLG}{LGL}}\right)*\frac{100}{150}\%} \quad \text{[Eq. 1]}$$

Students received automated feedback based on percentages derived from the formula as well as three levels (good, satisfactory, needs improvement). To validate the formula, 200 texts were randomly drawn from the sample. Two experts in scientific writing were blindly presented with these summaries and independently rated the summaries according to the summary criteria and three quality levels (good, satisfactory, needs improvement). The interrater reliability measured with Fleiss' kappa between the two experts was κ =.68, between rater 1 and the LSA-based summary scoring was κ =.73, between rater 2 and the LSA-based summary result was κ =.63, and between all three raters was κ =.68, which indicates substantial agreement (Landis & Koch, 1977; Seifried et al., 2012). Consequently, the formula's result satisfactorily represents the quality of the summaries as rated by humans.

4. The Present Research

We examined the effects of automated formative feedback on students' summarizing skills during six sessions implemented in an online university tutorial for undergraduate elementary education students. Reading scientific texts that address core aspects of teaching (e.g., teacher-student interaction) helps students to value scientific literature as a basis for their continuing development in linking theory and practice (Kunina-Habenicht, 2020). However, as shown above, students need support to create a mental model of a text, identify relevant aspects, and summarize effectively (Ahn, 2022; Duke & Pearson, 2009; Friend, 2001). Hence, we expected that providing automated formative feedback embedded in an online tutorial (*FALB*) will support the development of more sophisticated summarizing skills. Furthermore, we expected that the more frequent use of formative feedback will further positively impact summarizing skills. We tested our assumptions by examining the effectiveness of *FALB* with a group that regularly interacted with *FALB* (experimental group) compared to a control group. For this purpose, we formulated three research questions:

(RQ 1) Does the experimental group achieve a higher summary result than a control group?

(RQ 2) Which aspects of summarizing skills (content, avoidance of copied words, redundancy avoidance, relevance, and length) are particularly promoted by the automated formative feedback?

(RQ 3) Do students (experimental group only) who completed more feedback loops write higher quality summaries?

5. Methods

5.1. Participants

A total of 138 cases were included in this study, of which 87 students studied B.Ed. elementary school education (experimental group) and 51 participants studied M.A. special education at the same university (control group). In accordance with the educational curriculum of the bachelor's elementary school education program, participation in the online tutorial was mandatory for all elementary school education students and was worth one credit point. Data for the experimental group were collected in the summer semester 2019. Special education students participated voluntarily in the online tutorial and received one credit point in return. Data for the control group was collected in the summer semester of 2022 which was after three years of intensive online learning due to the COVID-19 pandemic. Therefore, the control group's performance might be at a higher level than if the data had been collected before the pandemic (see limitations).

The participating students were between 19 and 36 years old (M = 23.70, sd = 2.77) and were 84.1% female. 78.2% of the experimental group and 89.6% of the control group had not yet taken a class on teacher-student interaction in science education which was the topic of the texts to be summarized.

5.2. Procedure

As part of an online tutorial, all participants completed a demographic questionnaire and were informed about the criteria of summarizing used in this study. Participants of the experimental group received information on how to decode the automated feedback. Students had to summarize six texts, with two weeks intervals, using the computer-based learning environment *FALB*. In the first and last session, participants of the experimental group submitted their summary but did not receive feedback. For the other four texts, they could write their summary, upload it, and receive automated formative feedback up to ten times. Participants of the control group also had to summarize the same six texts, with one week interval and did not receive any feedback or comments on their summaries (see Appendix A for the curriculum of the tutorials). The difference between the completion times of the two groups had no pedagogical reason but was due to the curricula of the tutorials. However, we do not expect this difference to affect the results, as both groups were instructed to write the summary in 90 minutes without interruption. However, the experimental group could have suffered a slight disadvantage due to the two-week processing time.

The study was approved by the Institutional review board according to faculty regulations. The students provided informed consent for the use of their data. Confidentiality and personal data protection were guaranteed in accordance with relevant data privacy laws.

5.3. Data Analysis

Analysis was conducted using R (R Core Team, 2022), version 4.2.2, rStudio (Posit Team, 2022), the "psych" package (Revelle, 2022) for descriptive and correlational analyses, and the "lme4" (Bates et al., 2015) and "lmerTest" packages (Kuznetsova et al., 2017) to specify multilevel models of change. Tables were prepared with "apaTables" (Stanley, 2021) and models were drawn with "sjPlot" (Lüdecke et al., 2022). Cases with less than 10% content or more than 90% summary result at T0 were removed from the analysis because it either indicates the pretest was not summarized properly or students already possess skills to write high quality summaries (13 cases in the experimental group).

6. Results

Descriptive statistics for the summary result and the correlational analysis are shown in Table 1 and Table 2. The main interest of this study was to analyze students' improvement in summarizing across six time points when receiving automated formative feedback at four time points compared to a no treatment control group.

Table 2
Correlational Analysis of the Summary Result

Variable	1	2	3	4	5
1. Summary result_0					
2. Summary result_1	.19*				
3. Summary result_2	.00	.41**			
4. Summary result_3	-.10	.40**	.28**		
5. Summary result_4	.01	.34**	.49**	.39**	
6. Summary result_5	-.01	.27**	.35**	.32**	.22**

Note. * indicates $p < .05$. ** indicates $p < .01$.

First, we checked for a multilevel structure in the data by calculating the intraclass correlation. This revealed that 20.3% of the variance in the summary result over time was explained by individual differences justifying a second level. Hence, we used multilevel modeling of change with measurement points nested in students to account for interindividual as well as intraindividual change (Singer & Willett, 2003). To identify the optimal model, we tested several models with different functions of time as fixed and random effects using the deviance statistic (Table 3). If time is included as a fixed effect, the change in the dependent variable is set equal for all individuals. This implies that differences are estimated for individuals' intercepts, e.g., the summary result at T0 in this study, but not for the rate of change. If time is included as a random effect, both the intercepts and the change in the dependent variable can vary between individuals.

Table 3
Model Comparisons

Model	Test of deviance
No time – fixed time	$\chi^2 = 30.38$, $df = 1$, $p < .001$
Fixed time – fixed time^2	$\chi^2 = 13.59$, df = 1, p < .001
Fixed time^2 – fixed time^3	$\chi^2 = 40.24$, df = 1, p < .001
Fixed time^3 – fixed time^4	$\chi^2 = 3.71$, df = 1, p = .054
Fixed time^3 – random time	$\chi^2 = 9.99$, df = 1, p = .007
Fixed time^3 – random time^2	$\chi^2 = 26.47$, df = 1, p < .001
Fixed time^3 – random time^3	–

Note. Time^2 = quadratic change in time. Time^3 = cubic change in time. Time^4 = quartic change in time.

Table 1
Descriptive Statistics of the Summary Result over Time

			T0		T1		T2		T3		T4		T5	
Summary result	N	Min/Max	M	sd	M	sd	M	sd	M	sd	M	sd	M	sd
Control	51	10/100	50.65	15.40	53.77	16.57	49.57	17.13	47.98	15.29	46.92	14.97	53.06	15.06
Feedback	87		44.28	13.87	60.47	9.56	65.31	13.68	57.33	11.48	62.33	12.29	62.75	13.93

Tests of deviances showed that the summary result followed a cubic change (*time*^{3}) and the effects of time differed between individuals. The model with a fixed quartic slope (*time*^{4}) did not explain the data better than the model with a cubic change as a fixed effect ($\chi^2 = 3.71$, $df = 1$, $p = .054$). Thus, the fixed time^3 model was chosen the best and more parsimonious model. Next, we included *group* as a level-1 fixed effect to analyze different rates of change between the experimental group and the control group (Figure 4).

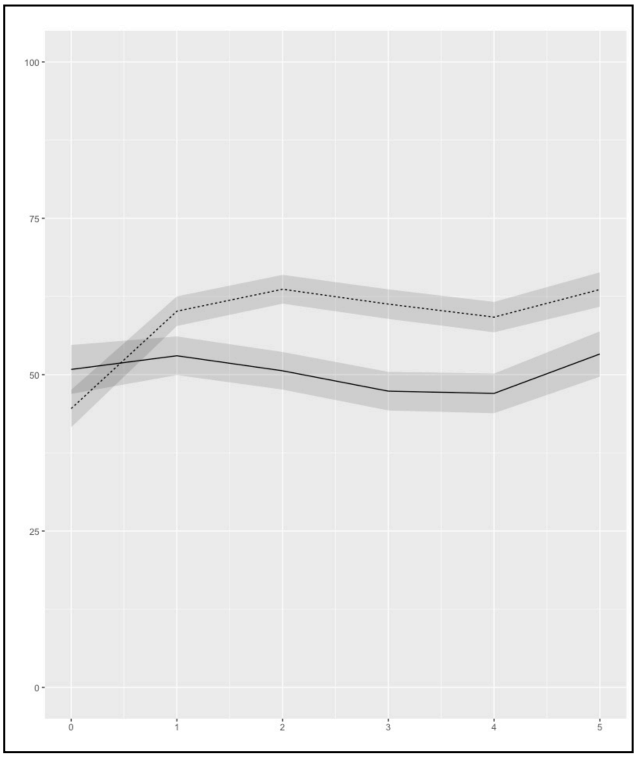


Figure 4

Predicted Values of Summary Result (y-axis) depending on Time (x-axis) for Control Group (solid) and Experimental Group (dashed)

The model estimates are presented in Table 4. The summary result at T0 differed significantly between the two groups ($\gamma_{\Delta\ control-feedback} = -6.26$, $p = .014$). The control group's summary result did not significantly change from T0 to T2, decreased slightly from T2 to T4, and then increased slightly once more from T4 to T5, thus following a cubic rate of change. On the contrary, the experimental group's summary result increased substantially from T0 to T2, decreased twice as much as the control group from T2 to T4, and then again increased slightly from T4 to T5 at the same rate of change as the control group. Thus, although the experimental group wrote significantly poorer summaries than the control group at T0, the experimental group benefited from the intervention and wrote significantly better summaries than the control group at T5 (RQ 1).

Table 4
Multilevel Model of Change in Summary Result

Fixed effects	*β*	*SE*	*p*
Time	0.64	3.28	.081
Time^2	-2.41	1.58	.009
Time^3	1.82	0.21	.003
Time ΔFB	1.99	4.13	.000
Time^2 ΔFB	-2.86	1.99	.013
Time^3 ΔFB	1.18	0.26	.121
Random effects	*Var*	*SD*	
Person	66.70	8.17	
Time	35.91	5.99	
Time^2	0.93	0.96	
Level-1 residual	141.54	11.90	
R^2_{total}	.12		

In a next step (RQ 2), we examined the single text quality scores (content, length, avoidance of copied words, redundancy avoidance; see appendix B for descriptive statistics, model comparisons, and estimates) to better understand which aspects of a summary were particularly promoted by the automated feedback (Figure 5). We omitted the relevance score from the analysis since more than 85.1% of all summaries had a relevance score over 90, indicating a ceiling effect.

Content followed a quadratic change (*time*^{2}), but the effects of time did not differ between individuals. Content was over 95% at T0 for both groups indicating a good content coverage. For the control group, the effects of time were not significant and content coverage remained at a high level. For the experimental group, content coverage decreased significantly from T0 to T3 and then increased from T3 to T5 ending at a level around 85%. *Length* followed a cubic change (*time*^{3}) and the effects of time differed between individuals. Overall, the length values ranged from just over 40% to just under 25%, slightly outside the intended range of 20 to 30% length of the original text. The control group's length was around 30% at T0. It increased slightly until T3 and decreased from T3 to T5. At T5, the mean length was a little less than 30%. Conversely, the experimental group's length started significantly higher at T0 and decreased until T5, ending with an average length below 25%. *Avoidance of copied* words followed a cubic change (*time*^{3}) and the effects of time differed between individuals. For the control group, the effects of time were not significant, and avoidance of copied words remained at the same level over time. By contrast, the experimental group reduced the copying of words significantly from T0 to T2, increased slightly from T2 to T4 and decreased again from T4 to T5.

Despite those variations, the level of avoidance of copied words always remained above 90 % from T2 to T5. *Redundancy avoidance* followed a cubic change (*time*^{3}), but the effects of time did not differ between individuals. Redundancy avoidance increased significantly from T0 to T2, decreased from T2 to T4, and then increased slightly from T4 to T5, all at a lower level than content and avoidance of copied words. No group differences were observed.

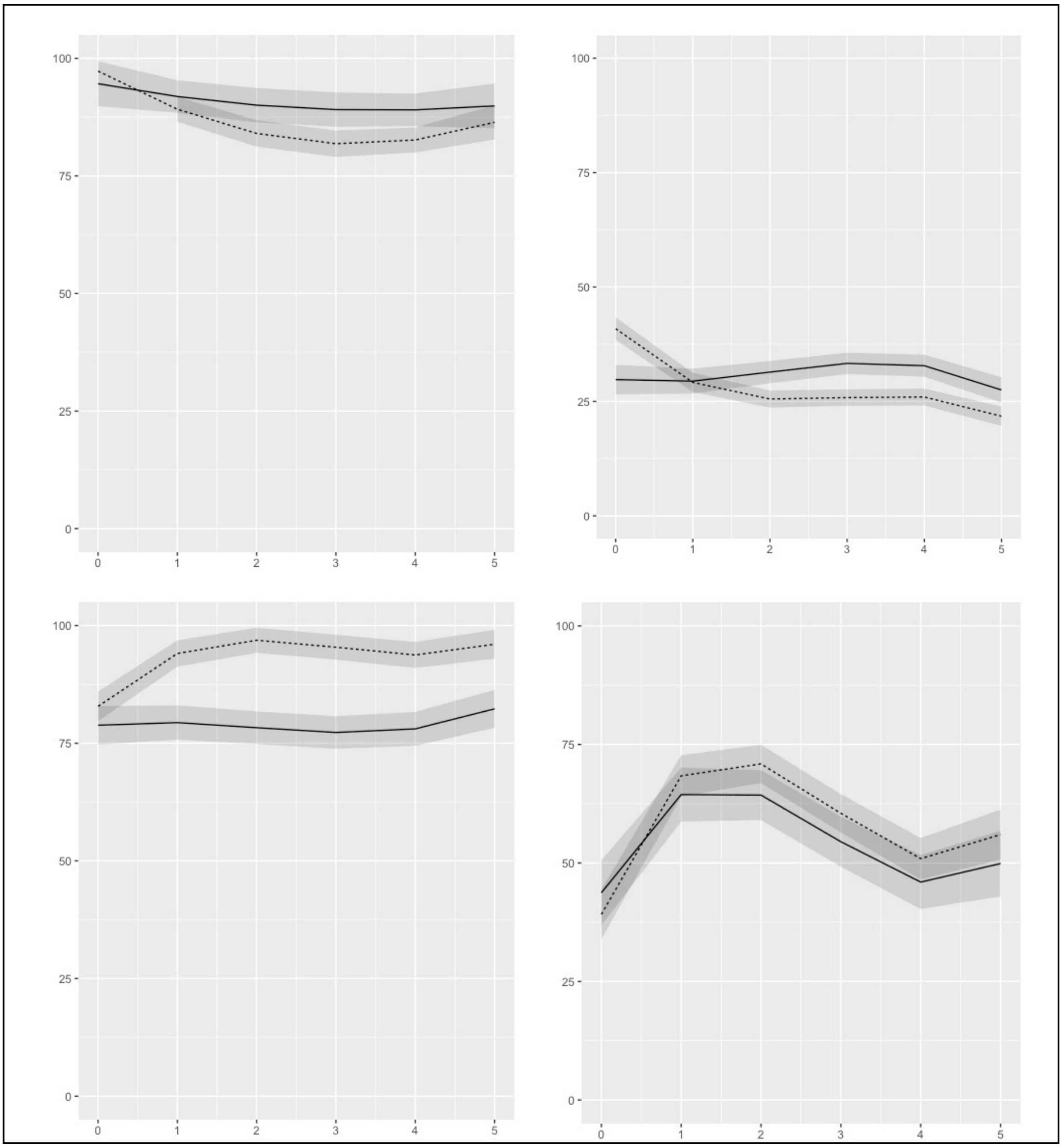


Figure 5

Predicted Profiles of Content (top-left), Length (top-right), Avoidance of Copied Words (bottom left), and Redundancy Avoidance (bottom right; all depended constructs on y-axis) depending on Time (x-axis) for Control group (solid) and Experimental group (dashed)

Last, we analyzed the experimental group individually regarding how frequently they used formative feedback (iteration) to improve the summary result (RQ 3; see appendix for descriptive statistics and correlations). The optimal model implied a small significant effect for the summary result on iteration, indicating that students who completed more feedback loops tended to write better summaries ($\beta = 0.14$, $p = .002$; Table 5). Taken together, the intervention with automated formative feedback supported students in writing better summaries, and specifically promoted adherence to the length requirement and avoidance of word copying. Furthermore, more feedback loops resulted in higher quality summaries.

Table 5
Fixed Effects of Summary Result with Time and Iteration as Independent Variables, Experimental Group only

Fixed effects	*β*	*SE*	*p*
Time	8.62	14.33	0.000
$Time^2$	-20.36	6.33	0.000
$Time^3$	13.74	0.84	0.000
Iteration	0.14	0.32	0.002
R^2_{total}	.34		

7. Discussion

The present study was conducted to examine the effectiveness of an online tutorial with automated formative feedback on promoting undergraduate elementary education students' improvement in summarizing. Summarizing skills are important to extract relevant information from scientific texts and to succeed in studying and graduating successfully. Our study showed that summarizing skills can be fostered by automated formative feedback (M. K. Kim & McCarthy, 2021; Wade-Stein & Kintsch, 2004).

First, the summary result was analyzed. The summary result is an overall evaluation of the summaries, including the five text quality scores and their relationship to each other. This allows comparing students' overall performance and inferring their improvement in summarizing skills. In the experimental group, during the four-session intervention, feedback highlighted the weaknesses in students' drafts and formatively verified their conformity to the summary criteria. The increase in the summary result for the experimental group from T0 to T1 indicates that the alignment of the internal and external reference standards was particularly strong. However, the summary result continued to be on a high level in the following sessions. In contrast, the summary result of the control group students did not change from T0 to T2, decreased slightly from T2 to T4, and then increased marginally from T4 to T5.

Thus, the formative feedback might have induced the students in the experimental group to regularly evaluate their drafts against the external reference standard and align the external to the internal reference standard (Narciss, 2017). Accordingly, the summaries submitted not only reflect the experimental group students' own ideas of what constitutes a good summary but are also the result of their understanding and internalization of the summary criteria. Furthermore, research has shown that task engagement declines over the course of a seminar (Darby et al., 2013). However, the feedback may have challenged experimental group students to consistently work at a higher level than control group students. This may suggest that not only the acquisition of effective summarizing skills was promoted, but also motivational regulation (Black & Wiliam, 2018; Clark, 2012). In the last session, students in the experimental group did not receive feedback while composing their summaries. However, the feedback group students' summary result remained at a significantly higher level than the control group students' summary result. We infer that the experimental group students may have transferred feedback insights while composing the final summary and thus achieved better summary results than the control group.

The single text quality scores demonstrate the interdependence of the summary criteria and provide more detailed information about how students have met each summary criterion over time. The longer and more redundant the summary, the higher the possibility of covering full content and avoiding copying words. At T0, students in both the feedback and control groups wrote summaries with a high content coverage, yet they copied more than 20% words from the original text, included more than 50% redundant passages and wrote summaries that were 30% (control group) or longer (experimental group) of the original text. This could indicate that students may not have effectively encoded and integrated textual information into their cognitive schema. In the tutorial, students were expected to write summaries that cover close to 100% of the content, while also scoring high on avoiding copied words and redundant passages as well as adhering to the 20-30% length limit. With this, we intended to stimulate deep processing of the original text and the development of a valid mental model. The profiles of the text quality scores illustrate the learning processes of the experimental group compared to the control group, which only had little rates of change and rather remained at baseline.

At T1, when students in the experimental group received formative feedback for the first time, they aligned their internal reference standard in terms of length and avoidance of copied words by submitting summaries within the optimal range of 20-30% of the original text and avoidance of copied words over 90%. However, the level of content coverage declined. From T2 to T3, while students maintained optimal levels of length and avoidance of copied words, content coverage continued to decrease slightly. Yet, in T4, content coverage increased again while the levels of length and avoidance of copied words remained in the optimal range. This suggests that students in the experimental group improved their summarizing skills over the four-session intervention by learning to coordinate summary criteria requirements and thus wrote short summaries in their own words while maintaining a high content coverage. These aspects address cognitive processes needed to create a mental model of a text (cf. Dunlosky et al., 2013; Friend, 2001; van Dijk & Kintsch, 1983). Consequently, it might be inferred that they learned to create more valid mental models of the texts. This assumption is further supported by the fact that students not only maintained an optimal length and few copied words throughout the intervention when they had to meet with the criteria, but also maintained these criteria on the final summary when they were not formatively monitored for adherence to the criteria or given formative feedback on their summaries.

The automated formative feedback could not foster redundancy avoidance. At T0, both the control and experimental groups started at a much lower level of redundancy avoidance than content and avoidance of copied words. This is partly because the summaries exceeded the length limit, increasing the risk of redundancy. Moreover, it might indicate insufficient skills in writing concisely. Additionally, students may have had unelaborated prior knowledge since 90% of all participants had not yet attended lessons about teacher-student interaction in science education (the topic of the original texts). With little prior knowledge, it is difficult to make inferences, condense the gist of a text, and reorganize its ideas, which hinders the ability to write concise summaries (K. Kim et al., 2019; van Dijk & Kintsch, 1983). From T1 to T5, group differences were not significant, and redundancy avoidance varied widely across texts, but remained at a lower level than content and avoidance of copied words. In the experimental group, the automated formative feedback fostered the revision phase in the writing process but did not explicitly address the planning phase. As a result, students may not have learned strategies to thoroughly condense the core aspects of the text and sufficiently restructure the ideas of the text – activities that often occur in the planning phase (Chew et al., 2019). Thus, the students were unable to condense the content into a concise summary and avoid redundancy at a higher level. For future designs, it would be worth investigating whether additional prompts that specifically promote the planning phase and activation of prior knowledge can help students develop strategies to better avoid redundancy. In addition, students may have had difficulty following the formative feedback on redundancy avoidance and understanding why some passages were marked as redundant. Thus, they may not have been able to benefit from the feedback.

Automated formative feedback can immediately provide valid feedback to almost an unlimited number of students (Lenhard, 2008; Seifried et al., 2012). In our study, more feedback loops (iterations) positively impacted the summary result. This observation supports findings from previous research which suggests that the frequency of revisions highly influences the quality of a summary (Kirkland & Saunders, 1991; Link et al., 2020; Roscoe et al., 2015; Zhu et al., 2020). Students in the experimental group who completed multiple feedback loops might have engaged more deeply in summarizing than students who sought less feedback (Zhang, 2020). These students may also have had more sophisticated skills in seeking and processing feedback (Narciss, 2017). They could have judged external feedback as relevant, understood it, and accepted it (Brown et al., 2016). Thus, they might have been willing to change their internal standards rather than reinterpret the automated feedback according to their internal standards (D. L. Butler & Winne, 2016). It would be beneficial to conduct a subsequent study to examine more closely how students engage with automated formative feedback.

8. Limitation

First, the data collection period of the two groups was far apart. Data collection for the feedback group was in 2019 and for the control group in 2022. Between these years, the COVID-19 pandemic occurred, which greatly changed teaching at universities by offering many courses online. Therefore, these groups are comparable to a limited extent, as the control group is more accustomed to a fully online tutorial. Compared to the students in the feedback group, the control group may have been less distracted from reading and writing the texts and their summaries on the computer. They may also have had more practice in summarizing because they may have had to document their work progress more frequently for other courses. This is also reflected in the control group's higher text quality at T0.

Second, the sample of the present study consists of solely elementary and special education students of one German university. Therefore, we do not know to what extent our results can be generalized to other university student populations and academic settings. For future studies the sample could be expanded to other educational programs and student populations to evaluate the generalizability of our findings beyond the scope of elementary teacher education. Furthermore, a longitudinal study over several semesters could provide insight into the long-term effects of automated formative feedback on students' academic growth and skill retention.

Third, *FALB* is a recently developed computer-based learning environment. Therefore, it has not yet been evaluated in terms of learning and feedback experience, and usability. Research has shown that satisfaction (Doménech-Betoret et al., 2017), feedback acceptance (Seifried et al., 2016), and technology acceptance (Hanham et al., 2021) are highly associated with learning success in computer-based learning environments. Thus, future studies should consider these moderating variables and how they affect feedback engagement and learning outcomes.

Fourth, this study lacks control variables like prior knowledge, language capability, and time on task. Research has shown that prior knowledge influences the creation of mental models (K. Kim et al., 2019; van Dijk & Kintsch, 1983), limited language capability may increase mental load, reduce reading comprehension, and shift attention (J. Li, 2014; McCutchen, 2011), and time on task is a strong predictor of text quality (J. A. Butler & Britt, 2010). Therefore, in future studies, such variables should be controlled to further explain interindividual differences in working with *FALB*.

9. Conclusion

Overall, this study provides valuable insight into the use of automated formative feedback through latent semantic analysis to teach effective summarizing skills to German university students. Furthermore, it highlights the necessity of imparting summarizing as an effective learning strategy and teaching effective summarizing strategies to students. We aimed to extend the evidence on automated summary evaluators through a new language, sample, and design. The findings demonstrated the potential of *FALB* to support the development of more sophisticated summarizing skills in German undergraduate elementary education students. They indicated that *FALB* particularly encouraged short summaries without copied words, with the formative nature of the feedback contributing to improved text quality. Yet, redundancy avoidance was promoted only to a limited extent, potentially attributable to the insufficient emphasis on the planning phase within the computer-based learning environment. Future research could advance these insights by investigating the efficacy of additional prompts during the planning phase in fostering enhanced redundancy avoidance and the adoption of effective summarizing strategies. Notably, the research by van den Boom et al. (2004, 2007) insinuates a complementary effect of prompting and feedback that remains unexplored in computer-based learning environments with automated formative feedback on summarizing. Taken together, the insights gained from this study on the support mechanisms of *FALB* are valuable and contribute to enhancing the design of intelligent feedback systems for university applications.

Acknowledgement

We express our gratitude to Prof. Dr. Wolfgang Lenhard for conText, his support in the development of *FALB*, and the enriching exchange. Additionally, we extend our thanks to Matthias Barde for the programming.

This project is part of the “Qualitaetsoffensive Lehrerbildung”, a joint initiative of the Federal Government and the Laender that aims to improve the quality of teacher training [grant number 01JA2016]. The program is funded by the Federal Ministry of Education and Research, Germany. The authors are responsible for the content of this publication.

References

Abba, K. A., Zhang, S. S., & Joshi, R. M. (2018). Community college writers' metaknowledge of effective writing. *Journal of Writing Research*, *10*(1), 85–105. https://doi.org/10.17239/jowr-2018.10.01.04

Ahn, S. (2022). Developing summary writing abilities of Korean EFL university students through teaching summarizing skills. *English Teaching*, *77*(2), 25–43. https://doi.org/10.15858/engtea.77.2.202206.25

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* (second edition, pp. 316–329). Guildford Press.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using **lme4**. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Beese, E. B. (2019). A process perspective on research and design issues in educational personalization. *Theory and Research in Education*, *17*(3), 253–279. https://doi.org/10.1177/1477878519893963

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, *21*(1), 5–31. https://doi.org/10.1007/s11092-008-9068-5

Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, *25*(6), 551–575. https://doi.org/10.1080/0969594X.2018.1441807

Brown, G. T. L., Peterson, E. R., & Yao, E. S. (2016). Student conceptions of feedback: Impact on self-regulation, self-efficacy, and academic achievement. *The British Journal of Educational Psychology*, *86*(4), 606–629. https://doi.org/10.1111/bjep.12126

Butler, D. L., & Winne, P. H. (2016). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, *65*(3), 245–281. https://doi.org/10.3102/00346543065003245

Butler, J. A., & Britt, M. A. (2010). Investigating instruction for improving revision of argumentative essays. *Written Communication*, *28*(1), 70–96. https://doi.org/10.1177/0741088310387891

Chew, C. S., Idris, N., Loh, E. F., Wu, W. V., Chua, Y. P., & Bimba, A. T. (2019). The effects of a theory-based summary writing tool on students' summary writing. *Journal of Computer Assisted Learning*, *35*(3), 435–449. https://doi.org/10.1111/jcal.12349

Clark, I. (2012). Formative assessment: Assessment is for self-regulated learning. *Educational Psychology Review*, *24*(2), 205–249. https://doi.org/10.1007/s10648-011-9191-6

Darby, A., Longmire-Avital, B., Chenault, J., & Haglund, M. (2013). Students' motivation in academic service-learning over the course of the semester. *College Student Journal*, *47*(1), 185–191.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, *41*(6), 391–407. https://doi.org/10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASI1>3.0.CO;2-9

Deeva, G., Bogdanova, D., Serral, E., Snoeck, M., & De Weerdt, J. (2021). A review of automated feedback systems for learners: Classification framework, challenges, and opportunities. *Computers & Education*, *162*, 104094. https://doi.org/10.1016/j.compedu.2020.104094

Doménech-Betoret, F., Abellán-Roselló, L., & Gómez-Artiga, A. (2017). Self-efficacy, satisfaction, and academic achievement: The mediator role of students' expectancy-value beliefs. *Frontiers in Psychology*, *8*, 1193. https://doi.org/10.3389/fpsyg.2017.01193

Duke, N. K., & Pearson, P. D. (2009). Effective practices for developing reading comprehension. *Journal of Education*, *189*(1–2), 107–122. https://doi.org/10.1177/0022057409189001-208

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

Foltz, P. W., Laham, D., & Landauer, T. K. (1999). Automated essay scoring: Applications to educational technology. *World Conference on Educational Multimedia, Hypermedia and Telecommunications*, 939–944.

Friend, R. (2001). Effects of strategy instruction on summary writing of college students. *Contemporary Educational Psychology*, *26*(1), 3–24. https://doi.org/10.1006/ceps.1999.1022

Garner, R. (1982). Efficient text summarization costs and benefits. *The Journal of Educational Research*, *75*(5), 275–279. https://doi.org/10.1080/00220671.1982.10885394

Graham, S. (2018). Instructional feedback in writing. In A. A. Lipnevich & J. K. Smith (Eds.), *The Cambridge Handbook of Instructional Feedback* (1st ed., pp. 145–168). Cambridge University Press. https://doi.org/10.1017/9781316832134.009

Hanham, J., Lee, C. B., & Teo, T. (2021). The influence of technology acceptance, academic self-efficacy, and gender on academic achievement through online tutoring. *Computers & Education*, *172*, 104252. https://doi.org/10.1016/j.compedu.2021.104252

Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487

Head, M. H., Readence, J. E., & Buss, R. R. (1989). An examination of summary writing as a measure of reading comprehension. *Reading Research and Instruction*, *28*(4), 1–11. https://doi.org/10.1080/19388078909557982

Hidi, S., & Anderson, V. (1986). Producing written summaries: Task demands, cognitive operations, and implications for instruction. *Review of Educational Research*, *56*(4), 473–493. https://doi.org/10.3102/00346543056004473

Hill, M. (1991). Writing summaries promotes thinking and learning across the curriculum—But why are they so difficult to write? *Journal of Reading*, *34*(7), 536–539. http://www.jstor.org/stable/40014578

Keck, C. (2006). The use of paraphrase in summary writing: A comparison of L1 and L2 writers. *Journal of Second Language Writing*, *15*(4), 261–278. https://doi.org/10.1016/j.jslw.2006.09.006

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

Kim, K., Clarianay, R. B., & Kim, Y. (2019). Automatic representation of knowledge structure: Enhancing learning through knowledge structure reflection in an online course. *Educational Technology Research and Development*, *67*(1), 105–122. https://doi.org/10.1007/s11423-018-9626-6

Kim, M. K., & McCarthy, K. S. (2021). Improving summary writing through formative feedback in a technology-enhanced learning environment. *Journal of Computer Assisted Learning*, *37*(3), 684–704. https://doi.org/10.1111/jcal.12516

Kirkland, M. R., & Saunders, M. A. P. (1991). Maximizing student performance in summary writing: Managing cognitive load. *TESOL Quarterly*, *25*(1), 105–121. https://doi.org/10.2307/3587030

Kunina-Habenicht, O. (2020). Wissen ist Macht: Ein Plädoyer für ein wissenschaftliches Lehramtsstudium. In C. Scheid & T. Wenzl (Eds.), *Wieviel Wissenschaft braucht die*

Lehrerbildung? (pp. 109–126). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-23244-3_6

Kürschner, C., & Schnotz, W. (2007). Konstruktion mentaler Repräsentationen bei der Verarbeitung von Text und Bildern. *Unterrichtswissenschaft*, *35*(1), 48–67. https://doi.org/10.25656/01:5486

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). **lmerTest** Package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). Introduction to latent semantic analysis. *Discourse Processes*, *25*(2 & 3), 259–284. https://doi.org/10.1080/01638539809545028

Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159–174.

Lee, A. V. Y., & Tan, S. C. (2017). Promising ideas for collective advancement of communal knowledge using temporal analytics and cluster analysis. *Journal of Learning Analytics*, *4*(3). https://doi.org/10.18608/jla.2017.43.5

Lemaire, B., & Dessus, P. (2001). A system to assess the semantic content of student essays. *Journal of Educational Computing Research*, *24*(3), 305–320. https://doi.org/10.2190/G649-0R9C-C021-P6X3

Lenhard, W. (2008). *Bridging the gap to natural language: A review on intelligent tutoring systems based on Latent Semantic Analysis*. https://opus.bibliothek.uni-wuerzburg.de/files/2397/Lenhard_Bridging_the_Gap.pdf

Lenhard, W., & Lenhard, A. (2014). *Berechnung des Lesbarkeitsindex LIX nach Björnson*. Unpublished. https://doi.org/10.13140/RG.2.1.1512.3447

Lenhard, W., Baier, H., Endlich, D., Lenhard, A., Schneider, W., & Hoffmann, J. (2012). Computerunterstützte Leseverständnisförderung: Die Effekte automatisch generierter Rückmeldungen. *Zeitschrift Für Pädagogische Psychologie*, *26*(2), 135–148. https://doi.org/10.1024/1010-0652/a000066

Lenhard, W., Baier, H., Hoffmann, J., & Schneider, W. (2007). Automatische Bewertung offener Antworten mittels Latenter Semantischer Analyse. *Diagnostica*, *53*(3), 155–165. https://doi.org/10.1026/0012-1924.53.3.155

Lenhard, W., Baier, H., Lenhard, A., Hoffmann, J., & Schneider, W. (2013). *ConText: Förderung des Leseverständnisses durch das Arbeiten mit Texten: Manual*. Hogrefe.

Li, H., Cai, Z., & Graesser, A. C. (2018). Computerized summary scoring: Crowdsourcing-based latent semantic analysis. *Behavior Research Methods*, *50*(5), 2144–2161. https://doi.org/10.3758/s13428-017-0982-7

Li, J. (2014). The role of reading and writing in summarization as an integrated task. *Language Testing in Asia*, *4*(1), 3. https://doi.org/10.1186/2229-0443-4-3

Link, S., Mehrzad, M., & Rahimi, M. (2020). Impact of automated writing evaluation on teacher feedback, student revision, and writing improvement. *Computer Assisted Language Learning*, *35*(4), 605–634. https://doi.org/10.1080/09588221.2020.1743323

Liu, M., Li, Y., Xu, W., & Liu, L. (2017). Automated essay feedback generation and its impact on revision. *IEEE Transactions on Learning Technologies*, *10*(4), 502–513. https://doi.org/10.1109/tlt.2016.2612659

Lüdecke, D. (2022). sjPlot: Data visualization for statistics in social science. R package version 2.8.12, https://CRAN.R-project.org/package=sjPlot

McCutchen, D. (2011). From novice to expert: Implications of language skills and writing-relevant knowledge for memory during the development of writing skill. *Journal of Writing Research*, *3*(1), 51–68.

McNamara, D. S., Roscoe, R., Allen, L., Balyan, R., & McCarthy, K. S. (2019). Literacy: From the perspective of text and discourse theory. *Journal of Language and Education*, *5*(3), 56–69. https://doi.org/10.17323/jle.2019.10196

Mok, W. S. Y., & Chan, W. W. L. (2016). How do tests and summary writing tasks enhance long-term retention of students with different levels of test anxiety? *Instructional Science*, *44*(6), 567–581. https://doi.org/10.1007/s11251-016-9393-x

Narciss, S. (2017). Conditions and effects of feedback viewed through the lens of the interactive tutoring feedback model. In D. Carless, S. M. Bridges, C. K. Y. Chan, & R. Glofcheski (Eds.), *Scaling up Assessment for Learning in Higher Education* (Vol. 5, pp. 173–189). Springer Singapore.

Nixon, S., Brooman, S., Murphy, B., & Fearon, D. (2016). Clarity, consistency, and communication: Using enhanced dialogue to create a course-based feedback strategy. *Assessment & Evaluation in Higher Education*, *42*(5), 812–822. https://doi.org/10.1080/02602938.2016.1195333

OpenAI, Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., Avila, R., Babuschkin, I., Balaji, S., Balcom, V., Baltescu, P., Bao, H., Bavarian, M., Belgum, J., … Zoph, B. (2023). *GPT-4 Technical Report*. https://doi.org/10.48550/ARXIV.2303.08774

Perin, D., Lauterbach, M., Raufman, J., & Kalamkarian, H. S. (2017). Text-based writing of low-skilled postsecondary students: Relation to comprehension, self-efficacy, and teacher judgments. *Reading and Writing*, *30*(4), 887–915. https://doi.org/10.1007/s11145-016-9706-0

Plass, J. L., & Pawar, S. (2020). Toward a taxonomy of adaptivity for learning. *Journal of Research on Technology in Education*, *52*(3), 275–300. https://doi.org/10.1080/15391523.2020.1719943

Posit Team (2022). RStudio: Integrated development environment for R. Posit Software, PBC, Boston, MA. http://www.posit.co/

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org

Revelle, W. (2022). Package "psych." *The Comprehensive R Archive Network*, *337*, 1–465.

Roscoe, R. D., Snow, E. L., Allen, L. K., & McNamara, D. S. (2015). Automated detection of essay revising patterns: Applications for intelligent feedback in a writing tutor. *Technology, Instruction, Cognition and Learning*, *10*, 59–79.

Schmid, R., Pauli, C., Stebler, R., Reusser, K., & Petko, D. (2022). Implementation of technology-supported personalized learning—Its impact on instructional quality. *The Journal of Educational Research*, 1–12. https://doi.org/10.1080/00220671.2022.2089086

Schnotz, W. (2006). Was geschieht im Kopf des Lesers? Mentale Konstruktionsprozesse beim Textverstehen aus der Sicht der Psychologie und der kognitiven Linguistik. *Text-Verstehen. Grammatik und darüber hinaus*, 222–238.

Schvaneveldt, R. W., & Cohen, T. A. (2010). Abductive reasoning and similarity: Some computational tools. In D. Ifenthaler, P. Pirnay-Dummer, & N. M. Seel (Eds.), *Computer-based Diagnostics and Systematic Analysis of Knowledge* (pp. 189–211). Boston, MA: Springer US.

Seifried, E., Lenhard, W., & Spinath, B. (2016). Automatic essay assessment: Effects on students' acceptance and on learning-related characteristics. *Psihologija*, *49*(4), 469–482. https://doi.org/10.2298/PSI1604469S

Seifried, E., Lenhard, W., Baier, H., & Spinath, B. (2012). On the reliability and validity of human and LSA-based evaluations of complex student-authored texts. *Journal of Educational Computing Research*, *47*(1), 67–92. https://doi.org/10.2190/EC.47.1.d

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153–189. https://doi.org/10.3102/0034654307313795

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford University Press.

Stanley, D. (2021). apaTables: Create American psychological association (APA) style tables. R package version 2.0.8, https://CRAN.R-project.org/package=apaTables

Stevens, E. A., Park, S., & Vaughn, S. (2019). A review of summarizing and main idea interventions for struggling readers in grades 3 through 12: 1978–2016. *Remedial and Special Education*, *40*(3), 131–149. https://doi.org/10.1177/0741932517749940

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 & Education*, *131*, 33–48. https://doi.org/10.1016/j.compedu.2018.12.005

Sung, Y.-T., Liao, C.-N., Chang, T.-H., Chen, C.-L., & Chang, K.-E. (2016). The effect of online summary assessment and feedback system on the summary writing on 6th graders: The LSA-based technique. *Computers & Education*, *95*, 1–18. https://doi.org/10.1016/j.compedu.2015.12.003

Sztajn, P., Confrey, J., Wilson, P. H., & Edgington, C. (2012). Learning trajectory based instruction: Toward a theory of teaching. *Educational Researcher*, *41*(5), 147–156. https://doi.org/10.3102/0013189X12442801

van den Boom, G., Paas, F. G. W. C., & van Merriënboer, J. J. G. (2007). Effects of elicited reflections combined with tutor or peer feedback on self-regulated learning and learning outcomes. *Learning and Instruction*, *17*(5), 532–548. https://doi.org/10.1016/j.learninstruc.2007.09.003

van den Boom, G., Paas, F., van Merriënboer, J. J. G., & van Gog, T. (2004). Reflection prompts and tutor feedback in a web-based learning environment: Effects on students' self-regulated learning competence. *Computers in Human Behavior*, *20*(4), 551–567. https://doi.org/10.1016/j.chb.2003.10.001

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 Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. Academic Press.

Wade-Stein, D., & Kintsch, E. (2004). Summary Street: Interactive computer support for writing. *Cognition and Instruction*, *22*(3), 333–362. https://doi.org/10.1207/s1532690xci2203_3

Westby, C., Culatta, B., Lawrence, B., & Hall-Kenyon, K. (2010). Summarizing expository texts. *Topics in Language Disorders*, *30*(4), 275–287. https://doi.org/10.1097/TLD.0b013e3181ff5a88

Zhang, Z. (2020). Engaging with automated writing evaluation (AWE) feedback on L2 writing: Student perceptions and revisions. *Assessing Writing*, *43*, 100439. https://doi.org/10.1016/j.asw.2019.100439

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

Appendices

Appendix A: Additional Information on FALB

Table A1
List of Text Material

Time	Text material	Word count	LIX
T0	Kleickmann, T.; Hardy, I.; Möller, K.; Pollmeier, J.; Tröbst, S. & Beinbrech, C. (2010). Die Modellierung naturwissenschaftlicher Kompetenz im Grundschulalter: Theoretische Konzeption und Testkonstruktion. Zeitschrift für Didaktik der Naturwissenschaften, 16; 268 – 269.	637	73,9
T1	Giest, H. (2015). Methodisches Erschließen. In: Kahlert, J.; Fölling-Albers, M.; Götz, M.; Hartinger, A.; Miller, S.; Wittkowske, S. (Hrsg.). Handbuch Didaktik des Sachunterrichts. Bad Heilbrunn: Verlag Julius Klinkhardt.	445	80,2
T2	Sodian B., Mayer D. (2013) Entwicklung des wissenschaftlichen Denkens im Vor- und Grundschulalter. In: Stamm M., Edelmann D. (eds) Handbuch frühkindliche Bildungsforschung. Springer VS, Wiesbaden.	531	66,0
T3	Leuchter, M. & Saalbach, H. (2014). Verbale Unterstützungsmaßnahmen im Rahmen eines naturwissenschaftlichen Lernangebots in Kindergarten und Grundschule. Unterrichtswissenschaft, 42(2), 117-131.	649	67,2
T4	Klieme, E.; Bürgermeister, A; Harks, B.; Blum, W.; Leiß. D & Rakoczy, K. (2010). Leistungsbeurteilung und Kompetenzmodellierung im Mathematikunterricht. Projekt Co2CA. Zeitschrift für Pädagogik, Beiheft; 56. Weinheim; Basel: Beltz.	618	74,6
T5	Möller, K., Steffensky, M. (2010). Naturwissenschaftliches Lernen im Unterricht mit 4- bis 8-jährigen Kindern. Kompetenzbereiche frühen naturwissenschaftlichen Lernens. In M. Leuchter (Ed.), Didaktik für die ersten Bildungsjahre. Unterricht mit 4- bis 8-jährigen Kindern. Seelze: Friedrich Verlag.	638	71,7

Table A2
Curriculum of the Experimental Group's Online Tutorial

Week	Assignment
1 + 2	Summarize Text 1 in FALB
3 + 4	Summarize Text 2 and interact with FALB
5 + 6	Summarize Text 3 and interact with FALB
7 + 8	Summarize Text 4 and interact with FALB
9 + 10	Summarize Text 5 and interact with FALB
11 + 12	Summarize Text 6 in FALB

Table A3
Curriculum of the Control Group's Online Tutorial

Week	Assignment
1	Summarize Text 1
2	Summarize Text 2
3	Summarize Text 3
4	Summarize Text 4
5	Summarize Text 5
6	Summarize Text 6
7	Recognizing persuasive arguments
8	Identifying persuasive argument structures
9	Write an argumentative essay in the field of pedagogy
10	Write an argumentative essay in the field of sustainability
11	Argumentation based on the Toulmin model
12	Recognizing and formulating persuasive arguments

Appendix B: Additional Information on RQ 2

Table B1.
Descriptive Statistics of Content, Length, Avoidance of Copied Words, Redundancy Avoidance

Variable	*min./max.*		T0		T1		T2		T3		T4		T5	
	0/100		*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
1. Content		Control	93.50	18.65	91.42	20.32	98.06	7.67	78.96	29.25	92.71	19.57	89.93	18.80
		Feedback	98.25	6.49	84.19	15.56	94.63	8.23	71.22	21.6	87.97	20.54	85.34	17.94
2. Length		Control	30.07	13.84	27.48	12.2	36.25	14.11	27.54	14.24	36.16	15.05	26.74	12.01
		Feedback	41.13	14.64	28.20	5.93	27.05	7.70	24.87	4.82	26.24	5.49	21.77	5.31
3. Avoid. cop. words		Control	78.47	19.31	81.07	18.33	75.04	20.66	80.42	21.97	76.54	18.96	82.59	20.80
		Feedback	82.51	16.15	95.41	6.90	94.70	8.51	96.97	5.28	93.25	10.08	96.07	7.49
4. Red. avoid		Control	42.16	25.67	68.95	29.56	61.74	24.04	50.60	27.75	51.14	26.77	47.96	29.39
		Feedback	38.05	19.80	70.56	24.17	73.16	23.66	51.61	24.77	58.65	24.19	53.77	24.84

Table B2
Model comparisons for Content

Model	Test of deviance
No time – fixed time	$\chi^2 = 24.29$, df = 1, p = .000
Fixed time – fixed time^2	$\chi^2 = 21.67$, df = 1, p = .000
Fixed time^2 – fixed time^3	$\chi^2 = 0.97$, df = 1, p = .324
Fixed time^2 – random time	-

Table B3
Model Comparisons for Length

Model	Test of deviance
No time – fixed time	$\chi^2 = 93.66$, df = 1, p = .000
Fixed time – fixed time^2	$\chi^2 = 7.88$, df = 1, p = .005
Fixed time^2 – fixed time^3	$\chi^2 = 36.68$, df = 1, p = .000
Fixed time^3 – random time	$\chi^2 = 35.68$, df = 1, p = .000
Fixed time^3 – random time^2	-

Table B4
Model Comparisons for Avoidance of Copied Words

Model	Test of deviance
No time – fixed time	$\chi^2 = 37.73$, df = 1, p = .000
Fixed time – fixed time^2	$\chi^2 = 14.31$, df = 1, p = .000
Fixed time^2 – fixed time^3	$\chi^2 = 26.42$, df = 1, p = .000
Fixed time^3 – random time	$\chi^2 = 19.18$, df = 1, p = .000
Fixed time^3 – random time^2	-

Table B5
Model Comparisons for Redundancy Avoidance

Model	Test of deviance
No time – fixed time	$\chi^2 = 0.00$, df = 1, p = .960
No time – fixed time^2	$\chi^2 = 61.97$, df = 1, p = .000
Fixed time^2 – fixed time^3	$\chi^2 = 75.00$, df = 1, p = .000
Fixed time^3 – random time	-

Table B6
Multilevel Model of Change in Content

Fixed effects	*Estimates*	*SE*	*p*
Intercept (control)	94.61	2.43	.000
Time (control)	-3.17	1.99	.112
$Time^2$ (control)	0.45	0.38	.245
Intercept ΔFB	2.68	3.06	.382
Time ΔFB	-6.42	2.51	.011
$Time^2$ ΔFB	1.04	0.48	.032
Random Effects	*Var*	*SD*	
Person	72.95	8.54	
Level-1 residual	278.72	16.70	

Table B7
Multilevel Model of Change in Length

Fixed effects	*Estimates*	*SE*	*p*
Intercept (control)	29.78	1.64	.000
Time (control)	-2.28	2.08	.272
$Time^2$ (control)	2.34	1.03	.023
$Time^3$ (control)	-0.39	0.13	.004
Intercept ΔFB	11.12	2.07	.000
Time ΔFB	-14.87	2.62	.000
$Time^2$ ΔFB	3.78	1.30	.004
$Time^3$ ΔFB	-0.30	0.17	.083
Random Effects	*Var*	*SD*	
Person	79.77	8.93	
Time	2.37	1.54	
Level-1 residual	60.20	7.76	

Table B8
Multilevel Model of Change in Avoidance of Copied Words

Fixed effects	*Estimates*	*SE*	*p*
Intercept (control)	78.82	2.09	.000
Time (control)	1.95	2.37	.410
$Time^2$ (control)	-1.68	1.17	.152
$Time^3$ (control)	0.29	0.15	.064
Intercept ΔFB	4.01	2.63	.127
Time ΔFB	14.85	2.99	.000
$Time^2$ ΔFB	-4.58	1.48	.002
$Time^3$ ΔFB	0.40	0.19	.039
Random Effects	*Var*	*SD*	
Person	147.01	12.13	
Time	3.34	1.83	
Level-1 residual	78.15	8.84	

Table B9

Multilevel Model of Change in Redundancy Avoidance

Fixed Effects	*Estimates*	*SE*	*p*
Intercept (control)	43.71	3.45	.000
Time (control)	34.82	6.04	.000
$Time^2$ (control)	-15.95	3.00	.000
$Time^3$ (control)	1.85	0.39	.000
Intercept ΔFB	-4.56	4.41	.302
Time ΔFB	12.37	7.60	.104
$Time^2$ ΔFB	-4.30	3.78	.255
$Time^3$ ΔFB	0.45	0.50	.364
Random Effects	*Var*	*SD*	
Person	132.37	11.51	
Level-1 residual	512.60	22.64	

Appendix C: Additional Information on RQ 3

Table C10

Means, Standard Deviations, and Correlations of Summary Result and Iteration

Variable	*M*	*SD*	1	2	3	4	5	6	7
1. tq_1	60.47	9.56							
2. tq_2	65.31	13.68	.46**						
3. tq_3	57.33	11.48	.24*	.10					
4. tq_4	62.33	12.29	.23*	.14	.26*				
5. iteration_1	2.39	1.77	.25*	.16	.23*	.28**			
6. iteration_2	2.53	2.08	.19	.18	.21	.05	.55**		
7. iteration_3	2.71	2.31	.23*	.11	.26*	.24*	.40**	.43**	
8. iteration_4	2.47	2.10	.24*	.33**	.20	.21	.48**	.43**	.31**

Note. * indicates $p < .05$. ** indicates $p < .01$.

Table C11
Multilevel Model of Change in Summary Result with Feedback Group only and Iteration as Independent Variable

Fixed effects	*Estimates*	*SE*	*p*
Time	72.08	14.33	.000
$Time^2$	-32.67	6.33	.000
$Time^3$	4.37	0.84	.000
Iteration	0.98	0.32	.000
Random Effects	*Var*	*SD*	
Person	23.62	4.86	
Level-1 residual	110.61	10.52	

Manuscript 2

Barkela, V., Schmitt, L., & Leuchter, M. (2023). The impact of cognitive and motivational resources on engagement with automated formative feedback. *Contemporary Educational Psychology*, 102234. https://doi.org/10.1016/j.cedpsych.2023.102234

Impact factor: 10.3

THE IMPACT OF COGNITIVE AND MOTIVATIONAL RESOURCES ON ENGAGEMENT WITH AUTOMATED FORMATIVE FEEDBACK

Veronika Barkela, Lukas Schmitt, & Miriam Leuchter

Abstract

The effectiveness of automated formative feedback highly depends on student feedback engagement that is largely determined by learners' cognitive and motivational resources. Yet, most studies have only investigated either cognitive resources (e.g., mental effort), or motivational resources (e.g., expectancy-value-cost variables). The purpose of this study is to examine the development (indicated by time) and relationship of 1) cognitive, 2) affective, and 3) behavioral feedback engagement as a function of cognitive and motivational resources in a computer-based learning environment with automated formative feedback. Data was collected from N = 330 German B.Ed. Elementary Education students who worked four consecutive sessions on summarizing texts. Previously invested mental effort (t-1) affected situational expectancy and cost but not situational value. 1) Cognitive feedback engagement was positively associated with previous performance but neither associated with cognitive nor motivational resources. 2) Affective feedback engagement was positively associated with intrinsic value and negatively associated with situational expectancies, invested mental effort and previous performance. 3) Behavioral feedback engagement was positively associated with situational expectancies and invested mental effort. This study contributes to the understanding of student's cognitive and motivational structures when engaging with automated formative feedback.

Keywords: feedback engagement, expectancy-value theory, mental effort, automated formative feedback, summarization, pre-service teacher students

1. Introduction

In higher education, summarizing skills are important to quickly extract and process relevant information from scientific texts (Kürschner et al., 2006). However, undergraduates struggle to grasp scientific texts in depth, distinguish important information from unimportant information, and link information to their prior knowledge (Kintsch, 1990). Consequently, they experience difficulties in summarizing scientific texts precisely and in their own words (Kim & McCarthy, 2021). One way to promote summarizing skills is to provide automated formative feedback which assesses linguistic features related to the properties of a summary at word, sentence, and document level using natural language processing (Kim & McCarthy, 2021; Wade-Stein & Kintsch, 2004). In the present study, for example, the system informs about content coverage, copied words, redundancies, irrelevant information, and adequate length, accessible in multiple iterations (Authors, in revision).

However, the effectiveness of such feedback highly depends on the individual engagement with feedback (Handley et al., 2011; Price et al., 2011). Especially in computer-based learning environments, it is the learners' responsibility to use feedback in a way that supports their learning processes (Ali et al., 2018; Winstone et al., 2017). Kinsey (2022) showed that individual engagement is largely determined by learners' cognitive and motivational resources.

Although a growing body of research has investigated associations between cognitive resources and learning engagement (Dong et al., 2020; Y. Liu & Sun, 2021) or motivational resources and learning engagement (Putwain et al., 2019; Sun & Rueda, 2012), research about the relationship between cognitive and motivational resources and feedback engagement is scarce (Han, 2017). For researching cognitive and motivational influences on feedback engagement we used two theoretical perspectives. 1) The amount of invested mental effort as a cognitive resource and, 2) expectancy-value theory (EVT) as a framework for motivational resources might explain decisions to engage with feedback. In addition, we consider time, as the invested mental effort and EVT variables vary across learning situations (Eccles & Wigfield, 2020; Paas et al., 2005). Therefore, the purpose of this study is to examine the relationships between mental effort and EVT variables with feedback engagement over multiple time points in computer-based learning environments with automated formative feedback on summarizing texts. Information about which resources influence feedback engagement may indicate ways to promote student feedback engagement.

1.1 Student Engagement with Automated Formative Feedback

The formative feedback process is a communicative act formed by feedback provider and recipient (Price et al., 2011). In computer-based learning environments, feedback is often computed by algorithms, hence on the part of the provider, the communicative act is limited. Therefore, the quality of the feedback communication processes depends highly on the recipients and their use of the feedback to its full potential (Handley et al., 2011; Van der Kleij & Lipnevich, 2021; Winstone et al., 2017). Following, the feedback recipients' individual levels of engagement determine if the feedback processes will be successful and support the achievement of the learning goals (Ali et al., 2018; Hyland & Zhang, 2018). Engagement is a multi-dimensional construct which encompasses cognitive, affective, and behavioral aspects (Ellis, 2010; Handley et al., 2011).

Cognitive feedback engagement describes the willingness to process and apply feedback to improve the outcome (Handley et al., 2011). Zhang and Hyland (2022) analyzed students processes of text revisions to examine cognitive engagement. They argue that cognitive engagement results in identifying errors and weaknesses and setting goals for improvement. These operations translate into improvement in text quality (Butler & Britt, 2010). Therefore, in this study, we chose the change in text quality from the initial summary to the finally submitted summary as an indicator of cognitive feedback engagement.

Affective feedback engagement relates to the perceived valence of the feedback (Mayordomo et al., 2022). Students' perception of the valence of the feedback determines the successful use of feedback and influence its effectiveness on learning (Seifried et al., 2016; Van der Kleij & Lipnevich, 2021). However, students often are discontent with the feedback they received (Adams et al., 2020; Zhang & Hyland, 2022). Especially when receiving automated feedback, students raise concerns about its accuracy and helpfulness (Lipnevich & Smith, 2009). Studies have shown that students who assumed they were receiving automated feedback accepted it to a lesser extent than students who assumed they were receiving feedback from a human, regardless of whether the feedback came from a human or was automated (Seifried et al., 2016). Therefore, in this study, we chose feedback acceptance as an indicator of affective feedback engagement.

Behavioral feedback engagement refers to the activity initiated by the recipient after having received feedback (Handley et al., 2011; Price et al., 2011). In the context of summarizing, this means that students with a high behavioral feedback engagement revise a text more often and process more feedback than students with low behavioral feedback engagement (M. Liu et al., 2017). Thus, in a computer-based learning environment that provides automated feedback, students with high behavioral feedback engagement tend to request newly calculated feedback more often (Van der Kleij & Lipnevich, 2021; Zhang & Hyland, 2022). Therefore, in this study, we chose the number of feedback iterations (the number of submissions of revised summaries and associated calculated feedback in one session) as an indicator of behavioral feedback engagement.

1.2 Cognitive Resources

Successfully engaging in complex tasks such as summarizing requires cognitive resources to process new information and integrate it into the individual's knowledge network (Sweller, 2020). Humans perceive mental effort as costly and tend to minimize its expenditure (Kool et al., 2017; Shenhav et al., 2017; Yee & Braver, 2018). Kool et al. (2010) showed that people tend to repeatedly choose actions that are associated with fewer cognitive demands. This is confirmed by Gieseler et al. (2020) who found that people choose less effortful task alternatives when exerting mental effort on an initial task. However, the prospect of reward and the expected efficacy of task performance may positively influence the willingness to invest mental effort (Frömer et al., 2021). Automated feedback and support in composing a summary might serve as such reward and affect efficacy judgements.

1.3 Motivational Resources

Building on EVT, expectancies, task values, and costs are key factors that influence decisions to engage in learning processes and to pursue activities that benefit learning (Eccles & Wigfield, 2020; Rosenzweig et al., 2019). Thus, in our case, we assume that students' expectancies, values, and costs influence their decision to engage with automated formative feedback in order to succeed in the task.

Expectancies. Expectancies for success refer to future performances through the anticipation of how well one will perform on an upcoming task. Expectancies thus express a situation-specific interpretation of one's own competence beliefs about the task (Eccles & Wigfield, 2020). Individuals' competence beliefs are derived from past performances. They are broad and stable subjective perceptions of respective current abilities (Marsh et al., 2012).

Task values. Task values describe subjective perceptions of the valence of a task and comprise the dimensions intrinsic value, utility value, attainment value, and costs. Subjective task values impact individual choices to engage with a learning activity (Rosenzweig et al., 2019). Intrinsic value comprises interest and enjoyment as well as willingness to engage in a task. Wigfield and Eccles (2020) link intrinsic value to the concepts of situational interest (Ryan & Deci, 2020) and intrinsic motivation (Hidi & Renninger, 2006) emphasizing the variability depending on task and time. Utility value denotes that the task is perceived as useful for goal achievement. According to Eccles and Wigfield (2020), utility value can be closely related to the concept of extrinsic motivation (Ryan & Deci, 2020). Attainment value refers to the personal value a student attributes to the task (Wigfield et al., 1997). Cost is conceptualized as perceived negative consequences of engaging with a task such as emotional distress and fear of missing opportunities. Eccles and Wigfield (2020) emphasize that every learning activity has costs, which are related to the benefits resulting in a cost to benefit ratio.

In EVT, the total value of a task is conceptualized as the sum of intrinsic value, utility value, attainment value, and cost (Eccles & Wigfield, 2020). Therefore, in many studies the value components were aggregated to a composite value scale (Alipio, 2020; Perez et al., 2014; Viljaranta et al., 2009; Wang & Eccles, 2013). In recent years, however, researchers suggested to exclude cost due to its negative valence (Barron & Hulleman, 2015; Flake et al., 2015; Jiang et al., 2018).

1.4 The Joint Impact of Cognitive and Motivational Resources on Feedback Engagement in the Perspective of Time

Feedback engagement might be related to the level of mental effort students invest in a summarizing task. Students with low engagement might expend few mental effort and process text and feedback superficially, resulting in little cognitive feedback engagement, in our case improvement of the summary, less affective feedback engagement, such as acceptance of feedback, and less behavioral feedback engagement, such as feedback iteration (Miller, 2015). The opposite applies to high engaged students. However, the willingness to allocate mental effort might also depend on students' motivational resources (Dunn et al., 2019; Feldon et al., 2019). For example, Putwain et al. (2019), Fan and Williams (2010), and Wang and Eccles (2013) showed in samples of primary and secondary school students that expectancies, values, and their interactions predict different manifestations of student engagement. Furthermore, Guo et al. (2016) investigated in a series of factor analyses the unique contributions of self-concept and the four value components of EVT. Their findings indicate that the value components, particularly intrinsic value and low cost, are important predictors of student-reported effort and teacher-reported engagement. Moreover, Goldstein (2006) conducted interviews with students about their decision to engage in feedback in an ESL writing course and found that motivation was an important factor in their decision to engage with feedback. Thus, in our study, different associations between the aspects of cognitive and motivational resources may lead to differences in cognitive (change in text quality), affective (feedback acceptance), and behavioral feedback engagement (feedback iteration).

In the perspective of time, research has shown that students' motivational states fluctuate from one learning situation to another (Dietrich et al., 2019; Eccles & Wigfield, 2020). According to the model of interest development (Hidi & Renninger, 2006), external influences, in our study e.g., the opportunity to learn about summarizing intensively, can create situational interest which can then be maintained for a longer period. However, several studies have shown that intrinsic value can decline over time during a university course (Darby et al., 2013; Seifried et al., 2016). Focusing on the reciprocal relationship between motivation and effort, Marsh et al. (2016) identified that prior effort had a negative effect on subsequent self-concept and prior self-concept had a positive effect on subsequent effort. Furthermore, Han (2017) found that learner beliefs, such as self-concept and success expectation, can change depending on the context, and that there is a mutually reinforcing relationship between learner beliefs and feedback engagement. Dietrich et al. (2017) showed that students invested more effort in situations where they had previously expected to be successful or valued the task highly. Yet, they also had higher expectancies and intrinsic value when they had invested more effort in the previous situation. A subsequent profile analysis showed that experienced costs remained fairly stable (Dietrich et al., 2019). This contradicts the findings of

Perez et al. (2019), which suggest less stability of the cost construct over time compared to the other task value constructs.

2. Rationale of Our Study

Based on the literature, motivational resources can be expected to influence cognitive, affective, and behavioral feedback engagement both directly and moderated by the invested mental effort (Paas et al., 2005; Shenhav et al., 2021). Yet, previously invested mental effort can also affect situational motivational resources which impact feedback engagement (Dietrich et al., 2017; Gieseler et al., 2020). Since previous performance has been found to be related to EVT variables and mental effort, it also ought to be considered as a control variable (Paas et al., 2005; Trautwein et al., 2009). From these theoretical considerations, a cross-lagged model can be assumed which describes the reciprocal relationship between cognitive and motivational resources and cognitive, affective, and behavioral engagement (Figure 1). To the best of our knowledge, there has been no previous research on the joint influence of cognitive and motivational resources on feedback engagement in a computer-based learning environment with automated feedback for summarizing.

3. Research Question and Hypotheses

With this study, we want to assess the explanatory power of the theoretical model in relation to our empirical data. Therefore, we specified the following research question and five hypotheses:

RQ: How do cognitive and motivational resources and previous performance affect feedback engagement over time?

We expect **direct** effects:

H1: Feedback engagement (cognitive, affective, behavioral) is positively affected by

- learner beliefs (self-concept, attainment/utility value, expectancy for success, intrinsic value),
- invested mental effort,

and negatively affected by

- cost.

Furthermore, we expect **indirect** paths:

H2: Invested mental effort is positively affected by

- learner beliefs (self-concept, attainment/utility value, expectancy for success, intrinsic value, cost).

Moreover, we expect **lagged** effects:

H3: Feedback engagement is positively affected by previous

- feedback engagement,
- performance.

H4: Learner beliefs (expectancy for success, intrinsic value, cost) are positively affected by previous

- feedback engagement,
- invested mental effort,
- performance.

H5: Invested mental effort is positively affected by previous

- feedback engagement,
- invested mental effort,
- performance.

4. Methods

4.1 Participants

In total, $N = 330$ German B.Ed. Elementary Education students participated in the study (274 female, 56 male). They were on average $M = 23.09$ ($SD = 2.89$) years old and at least in their fifth bachelor semester. In Germany, B.Ed. Elementary Education students take domain specific subjects in the first four semesters of their program and specialize in a particular type of school, in this case elementary school, in the last two semesters of their bachelor studies. 278 participants had not yet taken a course on teacher-student interaction which was the topic of the texts to be summarized. Data collection took place between April 2021 and December 2021.

The study was approved by the institutional review board according to faculty regulations. The students provided informed consent for the use of their data. Confidentiality and personal data protection were guaranteed in accordance with relevant data privacy laws.

4.2 Computer-based Learning Environment and Procedure

The *computer-based learning environment* is composed of a client side and a server side (see supplement, Figure A.1). The client side consists of a) the text to be summarized (Figure A.2), b) prompts encouraging the use of cognitive strategies for summarizing (Figure A.3), and c) stimuli for self-assessment presented once before students first access the feedback interface (Figure A.4). The feedback interface displays semantic feedback and score feedback. Semantic feedback provides information about copied words, redundancies, irrelevant sentences, and unknown words, which are marked in different colors in the text box (Figure A.5). Score feedback provides information about the summaries' length, how well the source text is covered, and how well copied words and repeating information are avoided (Figure A.6). The scores are saved as single scores and a composed score that is text quality. Semantic and score feedback can be obtained up to ten times.

The server's main task is to evaluate the summaries and to provide semantic feedback and score feedback. For this purpose, the source texts, expert summaries, and a semantic space were implemented on the server for calculation with latent semantic analysis. A more detailed description of the learning environment can be found in Authors (in revision).

Procedure. The intervention for summarizing was implemented in an online seminar about academic writing and lasted four weeks. Students had one week to complete each assignment within 90 minutes. Before the first session, students received video lectures on the required aspects of text quality and information about how to decode the automated feedback. In each session, students' task was summarizing a text. After receiving stimuli for self-assessment, they could revise the first draft, upload the revision, and receive automated formative feedback up to ten times. Before starting the intervention, the students were asked to complete a questionnaire about demographics, self-concept of summarizing, attainment value, and utility value. At each session, the students were asked about their expectancies for success, intrinsic value, and costs before starting the task and their invested mental effort and feedback acceptance after having finished the task.

4.3 Measures

The following variables were assessed once. *Self-concept of summarizing* focused on students' ability to summarize texts (Marsh et al., 2012). It was measured with five items on a four-point Likert-scale, e.g., 'Summarizing is one of my strengths' (Cronbach's $\alpha = .85$). *Summarizing utility value* measured the perceived utility of writing summaries in relation to studying successfully. The construct consisted of three items, e.g., 'I find summarizing very useful for my studies' (Cronbach's $\alpha = .76$). *Summarizing attainment value* was conceptualized as students' perceived value of summarizing to their studies. It was measured with five items, e.g., 'I find that summarizing is an important learning strategy' (Cronbach's $\alpha = .78$). In our study design, we incorporated findings from the literature, indicating that student teachers do not find theoretical knowledge (acquired through summarizing) that important (Bråten & Ferguson, 2015). Thus, we assumed that attainment and utility value would remain stable throughout the weeks and assessed them once before the course started.

The following variables were assessed in each session. *Cognitive feedback engagement* (change in text quality per session) is calculated by the difference of the last submitted summary and the first draft within the session. The range could theoretically span from -100 to +100. *Affective feedback engagement* (feedback acceptance) was measured with eight questions taken from Seifried et al. (2016), e.g., 'The feedback was fair'. It was measured on a four-point Likert-scale (1 = strongly disagree, 4 = strongly agree; Cronbach's $\alpha_{t1} = .94$, $\alpha_{t2} = .95$, $\alpha_{t3} = .96$, and $\alpha_{t4} = .97$). *Behavioral feedback engagement* (feedback iteration) is the number of how often the summary was resubmitted and feedback was requested. *Invested mental effort* was measured with three items based on Naismith et al. (2015) and Paas (1992), e.g., 'Today I had to concentrate very hard to understand the text' (Cronbach's $\alpha_{t1} = .72$, $\alpha_{t2} = .79$, $\alpha_{t3} = .71$, $\alpha_{t4} = .70$). *Expectancy for success was assessed* with one question about the students' belief how well their summary will score on a dimension with ten percent intervals (Doménech-Betoret et al., 2017).

Values about summarizing were assessed in line with conceptualizations by Wigfield et al. (1997) on a four-point Likert-scale (1 = strongly disagree, 4 = strongly agree). *Intrinsic value of summarizing* was constructed as the commitment on learning about summarizing, and comprised four items e.g., 'Today I like to summarize a text' (Cronbach's $\alpha_{t1} = .78$, $\alpha_{t2} = .83$, $\alpha_{t3} = .84$, and $\alpha_{t4} = .85$). *Perceived cost* of working with the learning environment was conceptualized as perceived emotional distress and fear of missing out on other opportunities when working with the learning environment (Eccles & Wigfield, 2020). It was measured with four items, e.g., ‘To write a good summary today, I must give up something I would have preferred to do right now’ (Cronbach’s α_{t1} = .72, α_{t2} = .80, α_{t3} = .81, and α_{t4} = .79). According to the literature (Hidi & Renninger, 2006), we assumed that intrinsic value and cost vary depending on task and time. We therefore assessed them before every session. *Previous performance* was measured as the summary’s final text quality score of the previous session. All self-report items are provided in the supplement B.

4.4 Data Analysis

We used the statistics program R, version 4.2.1, (R Core Team, 2022) for data analyses. An a priori power analysis was conducted using „semPower“ (Moshagen & Erdfelder, 2016) to determine the minimum sample size required to test the study hypotheses. Results indicated the required sample size to achieve 90% power for detecting a medium effect, at a significance criterion of α = .05, was N = 218 for structural equation modelling. Thus, the obtained sample size of N = 330 is a little overpowered but still adequate to test the study hypotheses.

Data processing and preparation was done using the R-package “psych” (Revelle, 2022). Confirmatory factor analyses and estimation of model fits were carried out using “lavaan” (Rosseel, 2012). The data were organized in a long format to represent one time point per subject in each row. Thus, the unit of analysis was each measurement occasion for each student to model the change in the variables over time. To create lagged variables, the variables’ values were shifted by one measurement point. This was done for the cognitive, affective, and behavioral feedback engagement and the invested mental effort variables.

5. Results

The main interest of this study was to examine the development and relationship of cognitive, affective, and behavioral feedback engagement as a function of cognitive and motivational resources, for which structural equation modeling is most appropriate. Descriptive statistics and correlational analysis are reported in supplement C. Informed by the literature (Eccles & Wigfield, 2020; Jiang et al., 2018; Viljaranta et al., 2009), and given the strong correlation in our study between summarizing utility and attainment value (r = .65; Table C.3), we created a composite value score with a high internal consistency (Cronbach's α = .85). Intrinsic value was measured at a situation-specific level and thus considered separately in the model.

5.1. Structural Equation Modeling

Following the theoretical assumptions, we specified and tested a cross-lagged model (Figure 2). *CFI*, *RMSEA,* and *SRMR* fell above or below the cut-offs (Hu & Bentler, 1999) and thus, indicated a good fit (CFI = .98 > .95; *TLI* = .96 > .95; RMSEA = .03 < .10; SRMR = .02 < .08). The chi-square/df-ratio did not exceed the cut-off factor 2 ($\chi^2(36) = 69.58$; χ^2/df = 1.93; Kyriazos, 2018).

5.2 Hypothesis Testing

We used structural equation modeling with the maximum likelihood estimator (ML) to test direct, indirect, and lagged relationships (Figure 2). All reported associations were significant at a p = .05 level (see also supplement D).

Direct Paths on feedback engagement (H1). Neither learner beliefs nor invested mental effort had a direct significant association with cognitive feedback engagement (change in text quality). Affective feedback engagement (feedback acceptance) had a negative association with expectancy for success (β = -.11) and a positive association with intrinsic value (β = .18). Furthermore, affective feedback engagement was negatively associated with invested mental effort (β = -.13). Behavioral feedback engagement (feedback iteration) was positively associated with expectancy for success (β = .14) and invested mental effort (β = .13). Cost had no direct effect on cognitive, affective, or behavioral feedback engagement.

Indirect Paths on feedback engagement (H2). We found positive associations between invested mental effort and intrinsic value (β = .11) and cost (β = .26). Furthermore, invested mental effort was negatively associated with self-concept (β = -.18). Besides, we found positive associations between self-concept and expectancy for success (β = .15) and attainment/utility value and intrinsic value (β = .31). Furthermore, attainment/utility value had a negative effect on cost (β = -.21).

Lagged Associations. Concerning H3, only previous performance (β = .11) had a positive effect on cognitive feedback engagement. Affective feedback engagement was negatively associated with previous performance (β = -.12) and positively associated with previous affective feedback engagement (β = .47). Behavioral feedback engagement was positively associated with previous behavioral feedback engagement (β = .39). Regarding H4, we found significant associations between expectancy for success and previous behavioral feedback engagement (β = .12), previously invested mental effort (β = -.14), and previous performance (β = .15). Furthermore, intrinsic value was positively associated with previous affective feedback engagement (β = .19) and cost was positively associated with previously invested mental effort (β = .23). Lastly, for H5, we found a positive association between invested mental effort and previous affective feedback engagement (β = .13). Besides, previously invested mental effort was significantly associated with previous cognitive feedback engagement (β = .13) and previous performance (β = -.17).

6. Discussion

This study investigated the joint impact of cognitive and motivational resources on the development of cognitive, affective, and behavioral feedback engagement in one model to draw conclusions on how to optimize a computer-based learning environment about summarizing and how to foster students' feedback engagement. Since studies have shown that invested mental effort is related to engagement (Dong et al., 2020; Y. Liu & Sun, 2021), and motivational variables can balance the costs to exert mental effort (Feldon et al., 2019; Yee & Braver, 2018), as well as reciprocal relationships between mental effort, motivational variables, and (feedback) engagement (Dietrich et al., 2017; Han, 2017; Marsh et al., 2016), we specified and tested a model with direct, indirect and lagged effects. Absolute model comparisons suggested a good fit implying that in this computer-based learning environment, students' feedback engagement is affected by motivational variables and invested mental effort. Moreover, situational motivational variables and invested mental effort are affected by previous feedback engagement and previous invested mental effort.

An integration of the three outcome variables in one model allows to account for the three dimensions of the engagement construct and their relation. Cognitive feedback engagement (change in text quality) correlates slightly positive with affective feedback engagement (feedback acceptance), suggesting that students who achieve higher change tend to better accept the feedback. Additionally, cognitive feedback engagement is moderately correlated with behavioral feedback engagement (feedback iteration), indicating that students revising their paper and receiving feedback more often are more likely to achieve higher change in text quality. Moreover, affective feedback engagement is negatively correlated with behavioral feedback engagement on a medium level, suggesting that students who accept their feedback tend to request less feedback. In the following, we will discuss the interplay of the cognitive and motivational resources with the three outcome variables.

Associations of T0 on T: In line with the theory, the positive relationship between self-concept and expectancy for success implies that students who think that they are good at summarizing expect high scores (Eccles & Wigfield, 2020). Self-concept influences negatively the invested mental effort, which is in line with previous research suggesting that students who believe to be capable of succeeding in a task experience less mental load than students who believe that they would not succeed in this task (K. M. Xu et al., 2021). Redifer et al. (2021) assume two causes for this relation. Students with high self-concept may a) shift attentional resources away from the demands of the task, thereby reducing their own cognitive load which results in less mental effort or b) experience a confidence booster and thus perceive the task as less difficult. Utility/attainment value is positively associated with intrinsic value and negatively associated with cost. This indicates that students who think that summarizing is important also experience high situational enthusiasm about writing a good summary and experience less emotional and opportunity costs (Perez et al., 2019). Unlike other studies (Guo et al., 2016; Meyer et al., 2019) utility/attainment value and self-concept are only slightly correlated, which may be attributable to the course topic (cf. Bråten & Ferguson, 2015).

Associations at T: Contrary to our hypothesis, expectancy for success is negatively associated with affective feedback engagement, indicating that students who expect lower summary scores are more likely to accept their feedback and are willing to improve their summaries according to the feedback. Students who expect higher summary scores might be disappointed by the feedback and therefore are less willing to accept it. Such behavior is known from research on self-efficacy and feedback acceptance. Within this framework, individuals tend to protect their self-efficacy beliefs when they receive feedback that is inconsistent with their efficacy judgments. They might question the accuracy of the feedback or attributing unsuccessful performance to bad luck (Nease et al., 1999; Silver et al., 1995). Furthermore, expectancy for success positively affects behavioral feedback engagement implying that students who expect to write a good summary tend to take more iterations (cf. Eccles & Wigfield, 2020; Putwain et al., 2019; Wu & Kang, 2021). However, there is a possibility that students who have high expectations but receive low scores do not engage thoroughly with the feedback but rather test the adaptivity of the automated feedback thus iterating more.

Moreover, the positive impact of intrinsic value on affective feedback engagement in the same learning situation illustrates that students who are more dedicated to writing a good summary attribute a high valence to their feedback. Additionally, intrinsic value is positively associated with the invested mental effort, indicating that students who are motivated to achieve higher scores also invest more mental effort. Cost is not directly associated with affective feedback engagement. However, cost is moderately associated with the invested mental effort. Thus, students might understand the investment of more mental effort as costs (Feldon et al., 2019). Invested mental effort positively impacts behavioral feedback engagement, indicating that students who allocated more cognitive resources requested feedback more often (Zhang & Hyland, 2022). On the contrary, invested mental effort is negatively associated with affective feedback engagement. Accordingly, students who experience higher mental effort are less likely to accept feedback. Thus, when having invested a high level of cognitive resources but not received the intended reward (good feedback), students may be disappointed and devalue the feedback, accepting it less (cf. Frömer et al., 2021).

Associations at T-1: Cognitive feedback engagement positively affects invested mental effort indicating that students who strongly improved their drafts invested more mental effort (cf. Sweller, 2020). Furthermore, performance negatively affects the invested mental effort (cf. Marsh et al., 2016).

Associations of T-1 on T: Previously invested mental effort is negatively associated with subsequent expectancy for success and positively associated with subsequent cost. This indicates that students who had allocated more cognitive resources earlier predicted a lower text quality in the subsequent learning situation. Effort is often considered a "double-edged sword" in previous research because it can affect students' self-perceptions, as having to invest more mental effort can imply less ability (Dietrich et al., 2017; Marsh et al., 2016). Furthermore, students who invested more mental effort in one learning situation experienced higher costs in the following session. This is in line with other studies which have shown that exerting cognitive control is effortful and therefore costly (Kool et al., 2017; Shenhav et al., 2021). The increase in students' costs depending on previous mental effort might be related to the finding that after having completed a high demanding task, people tend to choose less demanding tasks (Gieseler et al., 2020). In this study, however, the task difficulty remained comparably the same at each session and might thus have increased students' costs.

Previous performance positively affects expectancy for success. This finding is in line with EVT which states that expectancies for success build on experiences from past learning situations (Eccles & Wigfield, 2020). Furthermore, it positively impacts cognitive feedback engagement, indicating that students with a better summary are more likely to improve their summary in a subsequent learning situation (cf. Putwain et al., 2019). Moreover, students who wrote a good summary in a previous learning situation are less affectively engaged in the subsequent learning situation, indicating that they perceive the feedback as less informative or helpful. This can denote that students with high quality previous summaries might have more difficulty identifying areas for improvement than students with low previous summaries (J. Xu & Zhang, 2022). This might be reinforced by our tool. Students with high quality previous summaries received high score feedback and might not have seen much value in applying the feedback further. At last, previous affective feedback engagement highly affects subsequent affective feedback engagement. This indicates that once students appreciate the automated feedback to be fair and representative of their summary, they continue to do so throughout the course, employing more motivational and cognitive resources. This is in line with studies that show that the way students perceive the valence of feedback is a determining factor in the successful utilization of feedback (Van der Kleij & Lipnevich, 2021). Likewise, previous behavioral feedback engagement highly affected subsequent behavioral feedback engagement indicating that students who tend to iterate more often in one session maintain this behavior throughout the course (M. Liu et al., 2017).

6.1 Limitations

This study was conducted with an online computer-based learning environment implemented in a university course about scientific writing. Students had to work weekly assignments to pass the course. However, they were free to decide when and where to work on the assignments. Thus, we had little control whether the students followed the instructions and worked properly with the program. However, this shortage contributes to the ecological validity of this study, as online courses have become a common practice at universities. Yet, students might not have taken the tasks and feedback as seriously as they would have in a higher-stakes situation.

With our data, we were unable to significantly relate cognitive feedback engagement to cognitive or motivational resources, which may have been due to the supportive potential of the feedback algorithm. The automated feedback is calculated based on predefined criteria about content, copied words, redundancies, and length. It seems that students with poor previous performance could use this information to substantially improve their drafts and are thus stimulated in their zone of proximal development (Vygotsky, 1978). However, students who already wrote good summaries could not use the information they received from the feedback to further improve their summaries. The feedback algorithm was not able to develop higher-order criteria and therefore could not feed back more sophisticated information about summarizing. Hence, we were not able to support the higher achieving students in their zone of proximal development. For further research, the algorithm should be extended to include higher-order information, which could be adaptive in the way that only high-scoring students are shown this information. Thus, our study indicates that the development of a more intelligent algorithm that also detects errors in the structure and ideation of a summary itself may be worthwhile.

6.2 Conclusion

Automated formative feedback helps to overcome the dilemma between individual student support and limited resources in large university courses. However, learning successfully in such learning environments depends on the individual feedback engagement. We showed that motivation reduces cost as well as invested mental effort and thus increases behavioral feedback engagement. In addition, we showed that students who have high behavioral feedback engagement tend to have high cognitive feedback engagement. Consequently, when designing learning environments with automated formative feedback, the willingness to allocate more mental resources and to request feedback more often needs to be addressed to motivate students to engage more deeply in their learning processes. However, our study allowed for insights into the interplay between feedback engagement and cognitive and motivational aspects. These might be useful for further studies that consider e.g., students' beliefs about learning and transfer the findings to other computer-based learning environments.

Funding

This project is part of the "Qualitätsoffensive Lehrerbildung", a joint initiative of the Federal Government and the Länder that aims to improve the quality of teacher training [grant number 01JA2016]. The program is funded by the Federal Ministry of Education and Research, Germany. The authors are responsible for the content of this publication.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Appendix. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cedpsych.2023.102234.

References

Adams, A.-M., Wilson, H., Money, J., Palmer-Conn, S., & Fearn, J. (2020). Student engagement with feedback and attainment: The role of academic self-efficacy. *Assessment & Evaluation in Higher Education*, *45*(2), 317–329. https://doi.org/10.1080/02602938.2019.1640184

Ali, N., Ahmed, L., & Rose, S. (2018). Identifying predictors of students' perception of and engagement with assessment feedback. *Active Learning in Higher Education*, *19*(3), 239–251. https://doi.org/10.1177/1469787417735609

Alipio, M. (2020). Predicting academic performance of college freshmen in the philippines using psychological variables and expectancy-value beliefs to outcomes-based education: A path analysis. *Education & Administration*. https://doi.org/10.35542/osf.io/pra6z

Barron, K. E., & Hulleman, C. S. (2015). Expectancy-Value-Cost Model of Motivation. In *International Encyclopedia of the Social & Behavioral Sciences* (pp. 503–509). Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.26099-6

Bråten, I., & Ferguson, L. E. (2015). Beliefs about sources of knowledge predict motivation for learning in teacher education. *Teaching and Teacher Education*, *50*, 13–23. https://doi.org/10.1016/j.tate.2015.04.003

Butler, J. A., & Britt, M. A. (2010). Investigating instruction for improving revision of argumentative essays. *Written Communication*, *28*(1), 70–96. https://doi.org/10.1177/0741088310387891

Darby, A., Longmire-Avital, B., Chenault, J., & Haglund, M. (2013). Students' motivation in academic service-learning over the course of the semester. *College Student Journal*, *47*(1), 185–191.

Dietrich, J., Moeller, J., Guo, J., Viljaranta, J., & Kracke, B. (2019). In-the-moment profiles of expectancies, task values, and costs. *Frontiers in Psychology*, *10*, 1662. https://doi.org/10.3389/fpsyg.2019.01662

Dietrich, J., Viljaranta, J., Moeller, J., & Kracke, B. (2017). Situational expectancies and task values: Associations with students' effort. *Learning and Instruction*, *47*, 53–64. https://doi.org/10.1016/j.learninstruc.2016.10.009

Doménech-Betoret, F., Abellán-Roselló, L., & Gómez-Artiga, A. (2017). Self-efficacy, satisfaction, and academic achievement: The mediator role of students' expectancy-value beliefs. *Frontiers in Psychology*, *8*, 1193. https://doi.org/10.3389/fpsyg.2017.01193

Dong, A., Jong, M. S.-Y., & King, R. B. (2020). How does prior knowledge influence learning engagement? The mediating roles of cognitive load and help-seeking. *Frontiers in Psychology*, *11*, 591203. https://doi.org/10.3389/fpsyg.2020.591203

Dunn, T. L., Inzlicht, M., & Risko, E. F. (2019). Anticipating cognitive effort: Roles of perceived error-likelihood and time demands. *Psychological Research*, *83*(5), 1033–1056. https://doi.org/10.1007/s00426-017-0943-x

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

Ellis, R. (2010). Epilogue: A framework for investigating oral and written corrective feedback. *Studies in Second Language Acquisition*, *32*(2), 335–349. https://doi.org/10.1017/S0272263109990544

Fan, W., & Williams, C. M. (2010). The effects of parental involvement on students' academic self-efficacy, engagement and intrinsic motivation. *Educational Psychology*, *30*(1), 53–74. https://doi.org/10.1080/01443410903353302

Feldon, D. F., Callan, G., Juth, S., & Jeong, S. (2019). Cognitive load as motivational cost. *Educational Psychology Review*, *31*(2), 319–337. https://doi.org/10.1007/s10648-019-09464-6

Flake, J. K., Barron, K. E., Hulleman, C., McCoach, B. D., & Welsh, M. E. (2015). Measuring cost: The forgotten component of expectancy-value theory. *Contemporary Educational Psychology*, *41*, 232–244. https://doi.org/10.1016/j.cedpsych.2015.03.002

Frömer, R., Lin, H., Dean Wolf, C. K., Inzlicht, M., & Shenhav, A. (2021). Expectations of reward and efficacy guide cognitive control allocation. *Nature Communications*, *12*(1), 1030. https://doi.org/10.1038/s41467-021-21315-z

Gieseler, K., Inzlicht, M., & Friese, M. (2020). Do people avoid mental effort after facing a highly demanding task? *Journal of Experimental Social Psychology*, *90*, 104008. https://doi.org/10.1016/j.jesp.2020.104008

Goldstein, L. (2006). Feedback and revision in second language writing: Contextual, teacher, and student variables. In K. Hyland, & F. Hyland (Eds.), Feedback in second language writing: Contexts and issues (pp. 185e205). New York: Cambridge University Press.

Guo, J., Nagengast, B., Marsh, H. W., Kelava, A., Gaspard, H., Brandt, H., Cambria, J., Flunger, B., Dicke, A.-L., Häfner, I., Brisson, B., & Trautwein, U. (2016). Probing the unique contributions of self-concept, task values, and their interactions using multiple value facets and multiple academic outcomes. *AERA Open*, *2*(1), 233285841562688. https://doi.org/10.1177/2332858415626884

Han, Y. (2017). Mediating and being mediated: Learner beliefs and learner engagement with written corrective feedback. *System*, *69*, 133–142. https://doi.org/10.1016/j.system.2017.07.003

Handley, K., Price, M., & Millar, J. (2011). Beyond 'doing time': Investigating the concept of student engagement with feedback. *Oxford Review of Education*, *37*(4), 543–560. https://doi.org/10.1080/03054985.2011.604951

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist*, *41*(2), 111–127. https://doi.org/10.1207/s15326985ep4102_4

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Hyland, K., & Zhang, Z. V. (2018). Student engagement with teacher and automated feedback on L2 writing. *Assessing Writing*, *36*, 90–102. https://doi.org/10.1016/j.asw.2018.02.004

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

Kim, M. K., & McCarthy, K. S. (2021). Improving summary writing through formative feedback in a technology-enhanced learning environment. *Journal of Computer Assisted Learning*, *37*(3), 684–704. https://doi.org/10.1111/jcal.12516

Kinsey, A. W. (2022). *The relationship of cognitive and motivation as a predictor of persistence in accelerated online asynchronous courses* [Doctoral dissertation]. University of Memphis.

Kintsch, E. (1990). Macroprocesses and microprocesses in the development of summarization skill. *Cognition and Instruction*, *7*(3), 161–195. https://doi.org/10.1207/s1532690xci0703_1

Kool, W., McGuire, J. T., Rosen, Z. B., & Botvinick, M. M. (2010). Decision making and the avoidance of cognitive demand. *Journal of Experimental Psychology: General*, *139*(4), 665–682. https://doi.org/10.1037/a0020198

Kool, W., Shenhav, A., & Botvinick, M. M. (2017). Cognitive control as cost-benefit decision making. In T. Egner (Ed.), *The Wiley Handbook of Cognitive Control* (pp. 167–189). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118920497.ch10

Kürschner, C., Schnotz, W., & Eid, M. (2006). Konstruktion mentaler Repräsentationen beim Hör- und Leseverstehen. *Zeitschrift Für Medienpsychologie*, *18*(2), 48–59.

Kyriazos, T. A. (2018). Applied psychometrics: Sample size and sample power considerations in factor analysis (EFA, CFA) and SEM in general. *Psychology*, *09*(08), 2207–2230. https://doi.org/10.4236/psych.2018.98126

Lipnevich, A. A., & Smith, J. K. (2009). "I really need feedback to learn:" students' perspectives on the effectiveness of the differential feedback messages. *Educational Assessment, Evaluation and Accountability*, *21*(4), 347–367. https://doi.org/10.1007/s11092-009-9082-2

Liu, M., Li, Y., Xu, W., & Liu, L. (2017). Automated essay feedback generation and its impact on revision. *IEEE Transactions on Learning Technologies*, *10*(4), 502–513. https://doi.org/10.1109/tlt.2016.2612659

Liu, Y., & Sun, J. C.-Y. (2021). The mediation effects of task strategies on the relationship between engagement and cognitive load in a smart instant feedback system. *2021 International Conference on Advanced Learning Technologies (ICALT)*, 195–199. https://doi.org/10.1109/ICALT52272.2021.00065

Marsh, H. W., Pekrun, R., Lichtenfeld, S., Guo, J., Arens, A. K., & Murayama, K. (2016). Breaking the double-edged sword of effort/trying hard: Developmental equilibrium and longitudinal relations among effort, achievement, and academic self-concept. *Developmental Psychology*, *52*(8), 1273–1290. https://doi.org/10.1037/dev0000146

Marsh, H. W., Xu, M., & Martin, A. J. (2012). Self-concept: A synergy of theory, method, and application. In K. R. Harris, S. Graham, T. Urdan, C. B. McCormick, G. M. Sinatra, & J. Sweller (Eds.), *APA educational psychology handbook, Vol 1: Theories, constructs, and critical issues.* (pp. 427–458). American Psychological Association. https://doi.org/10.1037/13273-015

Mayordomo, R. M., Espasa, A., Guasch, T., & Martínez-Melo, M. (2022). Perception of online feedback and its impact on cognitive and emotional engagement with feedback. *Education and Information Technologies*, *27*, 7947–7971. https://doi.org/10.1007/s10639-022-10948-2

Meyer, J., Fleckenstein, J., & Köller, O. (2019). Expectancy value interactions and academic achievement: Differential relationships with achievement measures. *Contemporary Educational Psychology*, *58*, 58–74. https://doi.org/10.1016/j.cedpsych.2019.01.006

Miller, B. W. (2015). Using reading times and eye-movements to measure cognitive engagement. *Educational Psychologist*, *50*(1), 31–42. https://doi.org/10.1080/00461520.2015.1004068

Moshagen, M., & Erdfelder, E. (2016). A new strategy for testing structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*(1), 54–60. https://doi.org/10.1080/10705511.2014.950896

Naismith, L. M., Cheung, J. J. H., Ringsted, C., & Cavalcanti, R. B. (2015). Limitations of subjective cognitive load measures in simulation-based procedural training. *Medical Education*, *49*(8), 805–814. https://doi.org/10.1111/medu.12732

Nease, A. A., Mudgett, B. O., & Quinones, M. A. (1999). Relationships among feedback sign, self-efficacy, and acceptance of performance feedback. *Journal of Applied Psychology*, *84*(5), 806–814. https://doi.org/10.1037/0021-9010.84.5.806

Paas, F., Tuovinen, J. E., van Merriënboer, J. J. G., & Darabi, A. A. (2005). A motivational perspective on the relation between mental effort and performance: Optimizing learner involvement in instruction. *Educational Technology Research and Development*, *53*(3), 25–34. https://doi.org/10.1007/BF02504795

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

Perez, T., Cromley, J. G., & Kaplan, A. (2014). The role of identity development, values, and costs in college STEM retention. *Journal of Educational Psychology*, *106*(1), 315–329. https://doi.org/10.1037/a0034027

Perez, T., Dai, T., Kaplan, A., Cromley, J. G., Brooks, W. D., White, A. C., Mara, K. R., & Balsai, M. J. (2019). Interrelations among expectancies, task values, and perceived costs in undergraduate biology achievement. *Learning and Individual Differences*, *72*, 26–38. https://doi.org/10.1016/j.lindif.2019.04.001

Price, M., Handley, K., & Millar, J. (2011). Feedback: Focusing attention on engagement. *Studies in Higher Education*, *36*(8), Article 8. https://doi.org/10.1080/03075079.2010.483513

Putwain, D. W., Nicholson, L. J., Pekrun, R., Becker, S., & Symes, W. (2019). Expectancy of success, attainment value, engagement, and Achievement: A moderated mediation analysis. *Learning and Instruction*, *60*, 117–125. https://doi.org/10.1016/j.learninstruc.2018.11.005

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Redifer, J. L., Bae, C. L., & Zhao, Q. (2021). Self-efficacy and performance feedback: Impacts on cognitive load during creative thinking. *Learning and Instruction*, *71*, 101395. https://doi.org/10.1016/j.learninstruc.2020.101395

Revelle, W. (2022). Package "psych." *The Comprehensive R Archive Network*, *337*, 1–465.

Rosenzweig, E. Q., Wigfield, A., & Eccles, J. S. (2019). *Expectancy-value theory and its relevance for student motivation and learning* (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781316823279

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *American Statistical Association*, *48*(2), 1–36.

Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology*, *61*, 101860. https://doi.org/10.1016/j.cedpsych.2020.101860

Seifried, E., Lenhard, W., & Spinath, B. (2016). Automatic essay assessment: Effects on students' acceptance and on learning-related characteristics. *Psihologija*, *49*(4), 469–482. https://doi.org/10.2298/PSI1604469S

Shenhav, A., Fahey, M. P., & Grahek, I. (2021). Decomposing the motivation to exert mental effort. *Current Directions in Psychological Science*, *30*(4), 307–314. https://doi.org/10.31234/osf.io/yrd8n

Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., & Botvinick, M. M. (2017). Toward a rational and mechanistic account of mental effort. *Annual Review of Neuroscience*, *40*(1), 99–124. https://doi.org/10.1146/annurev-neuro-072116-031526

Silver, W. S., Mitchell, T. R., & Gist, M. E. (1995). Responses to successful and unsuccessful performance: The moderating effect of self-efficacy on the relationship between performance and attributions. *Organizational Behavior and Human Decision Processes*, *62*(3), 286–299. https://doi.org/10.1006/obhd.1995.1051

Sun, J. C.-Y., & Rueda, R. (2012). Situational interest, computer self-efficacy and self-regulation: Their impact on student engagement in distance education: Student engagement in distance education. *British Journal of Educational Technology*, *43*(2), 191–204. https://doi.org/10.1111/j.1467-8535.2010.01157.x

Sweller, J. (2020). Cognitive load theory and educational technology. *Educational Technology Research and Development*, *68*(1), 1–16. https://doi.org/10.1007/s11423-019-09701-3

Trautwein, U., Lüdtke, O., Roberts, B. W., Schnyder, I., & Niggli, A. (2009). Different forces, same consequence: Conscientiousness and competence beliefs are independent predictors of academic effort and achievement. *Journal of Personality and Social Psychology*, *97*(6), 1115–1128. https://doi.org/10.1037/a0017048

Van der Kleij, F. M., & Lipnevich, A. A. (2021). Student perceptions of assessment feedback: A critical scoping review and call for research. *Educational Assessment, Evaluation and Accountability*, *33*(2), 345–373. https://doi.org/10.1007/s11092-020-09331-x

Viljaranta, J., Nurmi, J.-E., Aunola, K., & Salmela-Aro, K. (2009). The role of task values in adolescents' educational tracks: A person-oriented approach. *Journal of Research on Adolescence*, *19*(4), 786–798. https://doi.org/10.1111/j.1532-7795.2009.00619.x

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Wade-Stein, D., & Kintsch, E. (2004). Summary Street: Interactive computer support for writing. *Cognition and Instruction*, *22*(3), 333–362. https://doi.org/10.1207/s1532690xci2203_3

Wang, M.-T., & Eccles, J. S. (2013). School context, achievement motivation, and academic engagement: A longitudinal study of school engagement using a multidimensional perspective. *Learning and Instruction*, *28*, 12–23. https://doi.org/10.1016/j.learninstruc.2013.04.002

Wigfield, A., Eccles, J. S., Yoon, K. S., Harold, R. D., Arbreton, A. J. A., Freedman-Doan, C., & Blumenfeld, P. C. (1997). Change in children's competence beliefs and subjective task values across the elementary school years: A 3-year study. *Journal of Educational Psychology*, *89*(3), 451–469. https://doi.org/10.1037/0022-0663.89.3.451

Winstone, N. E., Nash, R. A., Parker, M., & Rowntree, J. (2017). Supporting learners' agentic engagement with feedback: A systematic review and a taxonomy of recipience processes. *Educational Psychologist*, *52*(1), 17–37. https://doi.org/10.1080/00461520.2016.1207538

Wu, Y., & Kang, X. (2021). A moderated mediation model of expectancy-value interactions, engagement, and foreign language performance. *SAGE Open*, *11*(4), 215824402110591. https://doi.org/10.1177/21582440211059176

Xu, J., & Zhang, S. (2022). Understanding AWE feedback and english writing of learners with different proficiency levels in an EFL classroom: A sociocultural perspective.

The Asia-Pacific Education Researcher, *31*(4), 357–367. https://doi.org/10.1007/s40299-021-00577-7

Xu, K. M., Koorn, P., de Koning, B., Skuballa, I. T., Lin, L., Henderikx, M., Marsh, H. W., Sweller, J., & Paas, F. (2021). A growth mindset lowers perceived cognitive load and improves learning: Integrating motivation to cognitive load. *Journal of Educational Psychology*, *113*(6), 1177–1191. https://doi.org/10.1037/edu0000631

Yee, D. M., & Braver, T. S. (2018). Interactions of motivation and cognitive control. *Current Opinion in Behavioral Sciences*, *19*, 83–90. https://doi.org/10.1016/j.cobeha.2017.11.009

Zhang, Z., & Hyland, K. (2022). Fostering student engagement with feedback: An integrated approach. *Assessing Writing*, *51*, 100586. https://doi.org/10.1016/j.asw.2021.100586

Figure 1

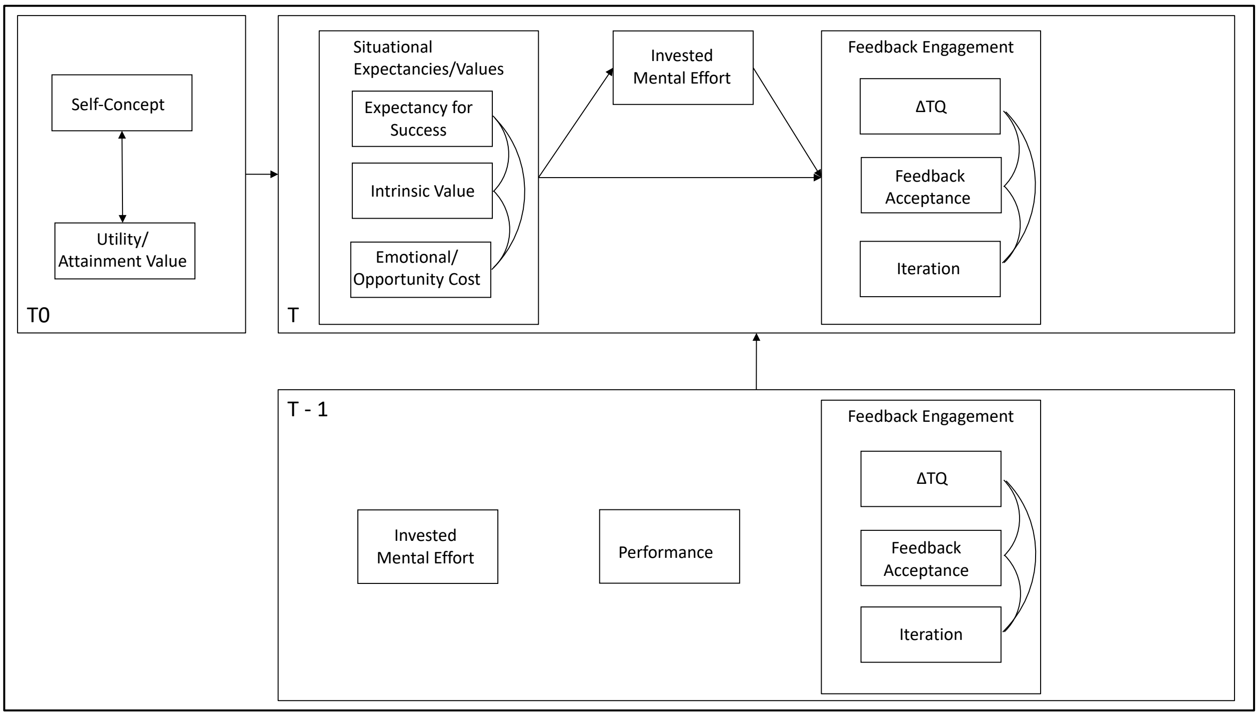


Fig. 1. Theoretical model describing reciprocal associations between invested mental effort, performance, motivational resources, and feedback engagement.

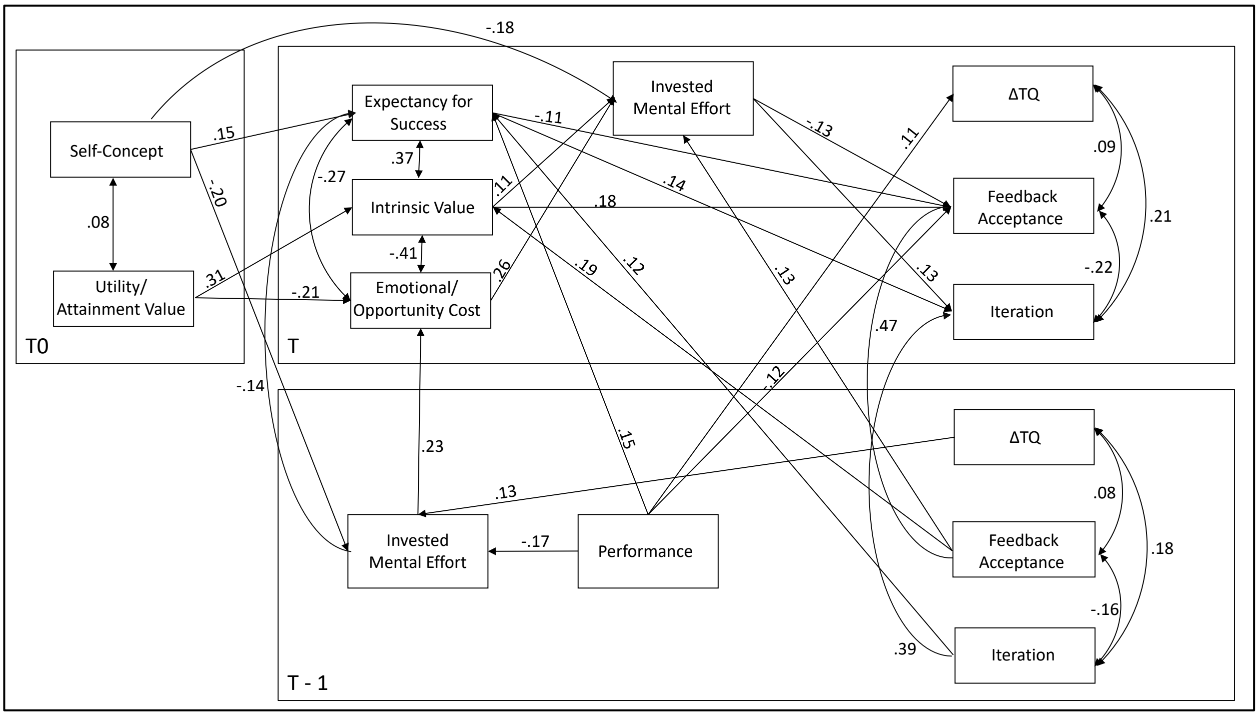


Fig. 2. Standardized path estimates and residual variances. Please note that only significant paths are printed, and all estimates are significant at a p = .05 level.

Manuscript 3

Barkela, V., & Leuchter, M. (2024). The interplay of instructional prompts and automated formative feedback. *Manuscript under revision for publication in the Journal of Computer Assisted Learning.*

Impact factor: 5.0

Honored with the Best Paper Award at EARLI-JURE 2021 (1st place).

The Interplay of Instructional Prompts and Automated Formative Feedback in Enhancing Summarizing Skills

Running title: Interplay of Prompts and Feedback

Veronika Barkela & Miriam Leuchter

Abstract

Background. Summarizing has been shown to be an effective learning strategy as it enhances cognitive skills, communication abilities, and information processing. Therefore, effective summarizing skills are an asset for university students. Yet, undergraduates struggle to employ effective summarizing strategies, thus needing support.

Objective. We aim at investigating the effectiveness of instructional prompts based on reciprocal teaching versus automated formative feedback in improving summarizing skills. We also explore if combining both prompts and feedback offers any additional benefits in enhancing these skills.

Methods. N = 254 elementary education students were asked to summarize six scientific texts, with one group (N = 75) receiving instructional prompts based on reciprocal teaching, one group (N = 87) receiving automated formative feedback, and one group (N = 92) receiving a combination of both.

Results and Conclusion. Instructional prompts based on reciprocal teaching alone support the development of summarizing skills, albeit to a lesser extent than automated formative feedback. Combining the two did not have a significant additional effect. At first sight, this may suggest a ceiling effect, indicating that learning potential is maximized with automated formative feedback. However, students in the combination group requested slightly less formative feedback, which indicates a possible substitution effect of automated formative feedback and instructional prompts based on reciprocal teaching.

Lay summary:

Automated feedback supports students to enhance their summarizing skills. Prompts support the enhancement of summarizing skills to a lesser extent. Combining the two is similarly effective to automated feedback alone.

Keywords: formative feedback, instructional prompts, reciprocal teaching, automated summary evaluator, summary writing, technology-enhanced learning environments, pre-service teacher students

1 Introduction

Summarizing has been shown to be an effective learning strategy as it enhances cognitive abilities, communication skills, and information processing (Dunlosky et al., 2013; Leopold et al., 2019; Stevens et al., 2019). Therefore, effective summarizing skills are a valuable asset for university students. However, undergraduate students often face challenges in understanding scientific texts in depth, discerning important information from less important details, and relating the information to their existing knowledge (Duke & Pearson, 2009; Friend, 2001; Green & Holman, 2021). Consequently, they struggle to write concise summaries using their own words (M. K. Kim & McCarthy, 2021). Therefore, learning support on text comprehension and summarizing skills is needed (Perin, 2019; Shanahan, 2019).

To foster reading and writing skills as means for summarizing, research has emphasized the important role of individualized formative feedback (Graham, 2018; Schunk & Rice, 1991). In large classes, however, such feedback is often neglected due to limited resources. Therefore, computer-based learning environments with automated formative feedback have been developed to support students' text comprehension and summarizing skills (Chew et al., 2019; M. K. Kim & McCarthy, 2021; Sung et al., 2016; Wade-Stein & Kintsch, 2004). In a previous study, we showed that students who practiced summarizing and received automated formative feedback improved their summarizing skills compared to a non-treatment control group that did not improve at all (Author, 2024). Furthermore, our results showed that students achieving higher text quality tended to engage in more feedback loops per session. The formative feedback focused on the writing phase, with students receiving feedback as they wrote. However, we did not encourage students to reflect on their own text comprehension, including strategies like assessing their understanding of key aspects of the texts or integrating new information with their existing knowledge before they started their writing process. Yet, prompting these strategies could enhance students' proficiency in text comprehension (Leopold et al., 2019), guiding them to employ more effective summarizing strategies and reducing reliance on automated formative feedback.

Prompts can be effective means in computer-based learning environments to support students in their learning (Hattie et al., 1996; Zheng, 2016), writing (Proske et al., 2012; Zellermayer et al., 1991) and summarizing (Ahn, 2022; Green & Holman, 2021). However, the prompts generally vary considerably, ranging from promoting students' self-regulation and metacognition (Engelmann & Bannert, 2021; Lim et al., 2023; Teng, 2022), stimulating reflection on the learning process and learning outcomes (Bannert, 2006; Krause & Stark, 2010), to guiding strategy use during the task processing (Lehmann et al., 2019; Proske et al., 2012). In this study, we developed a series of instructional prompts which explicitly stimulate reflection on text comprehension before starting to summarize, designed to guide the cognitive processes involved in summarizing and applying strategies accordingly.

Research by van den Boom (2004, 2007) indicates that the combination of instructional prompts and formative feedback supports learning more effectively than instructional prompts alone. However, in their study, prompts and feedback were provided by peers or human tutors. To the best of our knowledge, it has not yet been investigated whether these findings also apply to instructional prompts and formative feedback provided in a computer-based learning environment. Furthermore, it remains unclear whether the combination of instructional prompts and automated formative feedback supports the learning activities more effectively than automated formative feedback alone. Therefore, our study aims to investigate the interplay between instructional prompts and automated formative feedback and their effects on students' summarizing skills. To this end, we will investigate whether the instructional prompts we developed can support students' improvement in their summarizing skills compared to a pre-test. Additionally, we will compare this improvement to a feedback group receiving automated formative feedback. We also aim to determine whether the combination of prompts and automated formative feedback provides additional benefits to the learning process compared to receiving prompts alone or feedback alone. Finally, we will investigate whether the request for formative feedback can be manipulated by providing additional prompts.

1.1 Cognitive Processes in Summarizing

A summary is an abbreviated, concise version of a text that retains the most important aspects and necessary information from the original text while eliminating irrelevant and redundant parts (Dunlosky et al., 2013; Kirkland & Saunders, 1991; McAnulty, 1981). The quality of a summary is related to a person's mental model of the original text (M. K. Kim & McCarthy, 2021; Schnotz, 2006). Mental models are representations of a text and include not only explicitly stated information from the text but also inferences drawn from the text (van Dijk & Kintsch, 1983).

Summarizing entails multiple cognitive processes including activating prior knowledge, determining the relevance of information for inclusion in the summary, aligning the information given by the text with one's own prior knowledge, integrating new information into one's cognitive schema, constructing a mental model of the text, and translating this mental model into written text using one's own words (Becker-Mrotzek et al., 2014; Hidi & Anderson, 1986; Kellogg & Raulerson, 2007; Perin et al., 2017; Wade-Stein & Kintsch, 2004; Westby et al., 2010). People are better able to build comprehensive mental models and write good summaries when they have more corresponding prior knowledge and better coordination of these cognitive processes (Hathorn & Rawson, 2012; K. Kim et al., 2019). Because of this complex interplay of cognitive efforts, summarizing fosters text comprehension (Kintsch et al., 2000; Lenhard et al., 2012) as well as text composition (Graham, 2006; Hill, 1991; Kintsch, 1990). For assimilating information from scientific texts and summarizing appropriately, university students thus need advanced competencies that enable them to intentionally plan, monitor, and evaluate their writing processes (Zimmerman & Kitsantas, 2007). However, they differ in their prior knowledge and their skills to coordinate the processes and therefore might benefit from a learning environment that implements support to foster these competencies and teaches effective summarizing strategies.

1.2 Instructional Prompts for Fostering Effective Summarizing Strategies

For students with lower competencies, prompts can activate the use of effective summarizing strategies thus compensating for the lack of ability. Prompts are guidelines or questions directing students to intentionally use more sophisticated strategies which in turn foster their learning (Bannert, 2006). For example, Lehmann et al. (2019) reported positive effects for prompts that encouraged identifying a text's key elements and synthesizing ideas from various texts, which improved writing task performance. Engelmann et al. (2021) showed that prompts stimulated students to devote more time to relevant texts, which resulted in better learning compared to a control group. Moreover, they revealed that those students who interpreted the prompts as a request to reflect on their metacognitive learning activities performed better in a transfer test than those students who used the prompts as a call to action without deeper reflection on learning activities.

Being guided by the cognitive processes of summarizing, prompts should address the activation of prior knowledge, the assimilation of new information into the cognitive schema, and the integration of text content into the broader field. To this aim, the framework of reciprocal teaching that encompasses the three key strategies questioning, clarifying, and predicting (Brown et al., 1984) can provide parameters for formulating prompts to foster summarizing skills. Reciprocal teaching was developed to foster text comprehension in elementary and middle school but has been widely used and adapted in literacy education for all ages (Tolosa et al., 2015; Topping et al., 2013). *Questioning* involves formulating queries about the text to capture key aspects and information. With the aim of producing a coherent summary, questioning provides a context for text exploration and activation of prior knowledge which ensures comprehensive understanding. *Clarifying* comprises recognizing and elucidating unclear, challenging, or unfamiliar elements within a text. These aspects may involve awkward sentence structures, unknown vocabulary, unclear references, or complex concepts. For writing a clear summary, clarifying supports building a mental model of a text by engaging in activities such as re-reading, contextualizing the information in the text, and reflecting of the relationship between the theoretical aspects in the text and related practice (Fiorella & Mayer, 2016; Hmelo-Silver & Eberbach, 2012). *Predicting* entails the formulation of hypotheses about the text's direction and the author's purpose. For writing a comprehensive summary, these actions support the integration of new information in the reader's prior knowledge and promotes the understanding of the text's structure. Besides, predicting establishes a motive for reading, aiming to validate or challenge the hypotheses generated by the reader (Doolittle et al., 2006).

To the best of our knowledge, the potential of these strategies to effectively elicit summarizing strategies within a personalized computer-based learning environment for university students remains unexplored. Therefore, one aim of this study is to examine the support potential of instructional prompts following the reciprocal teaching framework. We are also interested in the interplay between these instructional prompts and automated formative feedback.

1.3 Automated Formative Feedback for Summarizing

External formative feedback involves aligning internal representations and an external standard of a good summary with the help of a feedback provider (cf. Black & Wiliam, 2009; Narciss, 2017). Key principles contributing to effective feedback include immediacy, elaboration, and individualization (Hattie & Timperley, 2007; Nixon et al., 2016; Shute, 2008). However, the challenge of delivering personalized support to students in large university classes, often constrained by limited resources, is a significant concern for educators (Allen et al., 2016). Automated formative feedback addresses this by enabling immediate and frequent assessment of students while adhering to feedback principles (Deeva et al., 2021; Strobl et al., 2019). It can tailor feedback to individual student work, mark problematic aspects, and offer suggestions for improvement. Furthermore, computer-based learning environments with automated feedback can save information, assessments, and completed assignments, providing convenient access for students to assignment objectives, criteria, previous work, and feedback. Additionally, they consistently evaluate texts using predefined criteria, ensuring reliability, with studies supporting their comparable validity to human feedback (Seifried et al., 2012; Sung et al., 2016).

Research has shown that providing automated formative feedback fosters students' summarizing and writing. Kim & McCarthy (2021) identified positive effects of automated feedback on university students' summaries, contributing to the development of a more robust and cohesive knowledge structure. Sung et al. (2016) observed beneficial outcomes in sixth graders' summarizing skills through the formative provision of concept maps and semantic feedback. Similarly, Chew et al. (2019) introduced a computer-assisted learning environment for summarizing, incorporating concept mapping, worked examples, and automated feedback on summarizing strategies, demonstrating its efficacy among university students. Zhu et al. (2020) highlighted that context-dependent automated formative feedback yielded a more substantial impact on learning compared to context-independent feedback. Authors (Author, 2024) demonstrated the effectiveness of a computer-based learning environment with semantic formative feedback on summarizing. Students in their study improved their summaries, and more feedback loops were associated with higher text quality.

To support the alignment of an internal reference standard with an external standard, setting specific text quality criteria that are consistent with the automated feedback plays a pivotal role in fostering deep text processing and summarizing skills (cf. Hill, 1991; Wade-Stein & Kintsch, 2004). Underlining the importance of including all essential aspects of the original text in a summary promotes a comprehensive reproduction of the text. Emphasizing the significance of avoiding copied words encourages students to express their thoughts independently. Stressing redundancy avoidance as an important aspect of text quality contributes to the creation of brief and concise summaries that capture the essence of the content. Finally, limiting the length of a summary serves to condense content to key messages.

2 The Present Research

Research has shown that providing students with instructional prompts support summarizing (Ahn, 2022; Green & Holman, 2021). Automated formative feedback helps to improve students' summarizing skills as well (Author, 2024; Lenhard et al., 2012; Wade-Stein & Kintsch, 2004). However, since the interplay of instructional prompts based on reciprocal teaching and automated formative feedback is still unclear, we aim to investigate differences between them as well as their relationship. As there may be additional benefits to combining prompts and feedback, we consider it worthwhile to explore them. Acknowledging the established role of time on task as a robust predictor of engagement and learning gains, particularly in multimedia learning settings, time on task should be considered as a control variable (Butler & Britt, 2010; Lee, 2018; Manwaring et al., 2017; Zhang & Hyland, 2022). Given that writing good summaries relies on both reading comprehension and language use (S.-A. Kim, 2001; Perin et al., 2017), language proficiency ought to be included as a control variable. Additionally, controlling for text difficulty accounts for variations in texts being summarized. Based on these theoretical considerations, a pre-post-test study was conducted with four intervention points. The following research questions were investigated:

1) Does a combination of instructional prompts and automated formative feedback enhance the learning of summarizing skills compared to a prompts-only and a feedback-only group?
2) Can the provision of instructional prompts engage students in maintaining a higher rate of feedback loops throughout the intervention compared to a feedback-only group?

3 Methods

3.1 Participants

A total of 254 Elementary Education students participated in the study (210 female, 44 male). They were $M = 23.09$ ($SD = 2.64$) years old and at least in their fifth bachelor semester. The B.Ed. Elementary Education students take domain specific subjects in the first four semesters of their program and specialize in a particular type of school, in this case elementary school, in the last two semesters of their bachelor studies. 79.5% of the participants had not yet taken a course on teacher-student interaction which was the topic of the texts to be summarized. Language proficiency was $M = 3.49$ ($SD = 0.80$; *min.* = 1, *max*. = 5).

The study was approved by the institutional review board according to faculty regulations. The students provided informed consent for the use of their data. Confidentiality and personal data protection were guaranteed in accordance with legally valid data privacy laws.

3.2 Learning Environment and Procedure

The *computer-based learning environment FALB* is designed to provide opportunities for improving effective summarizing skills. It allows to implement instructional prompts and / or automated formative feedback and is composed of two main components, front end and server. The server incorporates source texts, expert summaries, and a semantic space for computation using Latent Semantic Analysis (Landauer, T. K. et al., 1998). It allows for an automated scoring of students' texts according to predefined standards of content, avoidance of copied words, redundancy avoidance, relevance, and length as well as measuring time on task. A more detailed description of the theoretical and technical aspects of the learning environment can be found in Authors (Author, 2024).

The front end is defined as the platform which the user sees and interacts with. After starting with a page presenting the text to be summarized, a text editor is introduced that allows users to compose and revise the summary. While composing the summary, students access unlimited re-reading of the original text via a button. In the following, the differences between the prompts and feedback condition are described.

Prompts condition. In the prompts condition, a series of instructional prompts are presented to encourage the reflection and use of summarizing strategies after the presentation of the text to be summarized and before being presented with the text editor to write down and revise the summary. These prompts were developed based on reciprocal teaching theory (Brown et al., 1984; Table 1).

Table 1
Series of Instructional Prompts

Prompt	Strategy
The title of the text is "xxx". Have you ever acquired knowledge on this topic? In what context? What content can you still remember? Find up to three questions that are answered in the text.	Questioning
Can you comprehend the line of thought in the text? Is there an aspect that is unclear to you? If yes, read the relevant passage again. Has the aspect now been clarified? Can you illustrate the content of the text with up to three examples of your own?	Clarifying
What further questions do you have on this topic?	Predicting

Feedback condition. In the feedback condition, the feedback interface based on 'conText' (Lenhard et al., 2013) and 'summary street' (Wade-Stein & Kintsch, 2004) is presented via a button after the submission of the texts. It displays semantic and score feedback (Figure 1). Semantic feedback highlights copied words, redundancies, irrelevant sentences, and unknown words using various colours in the text box. Score feedback assesses summary length, content coverage of the original text, relevance, as well as avoidance of copied words and repeated information.

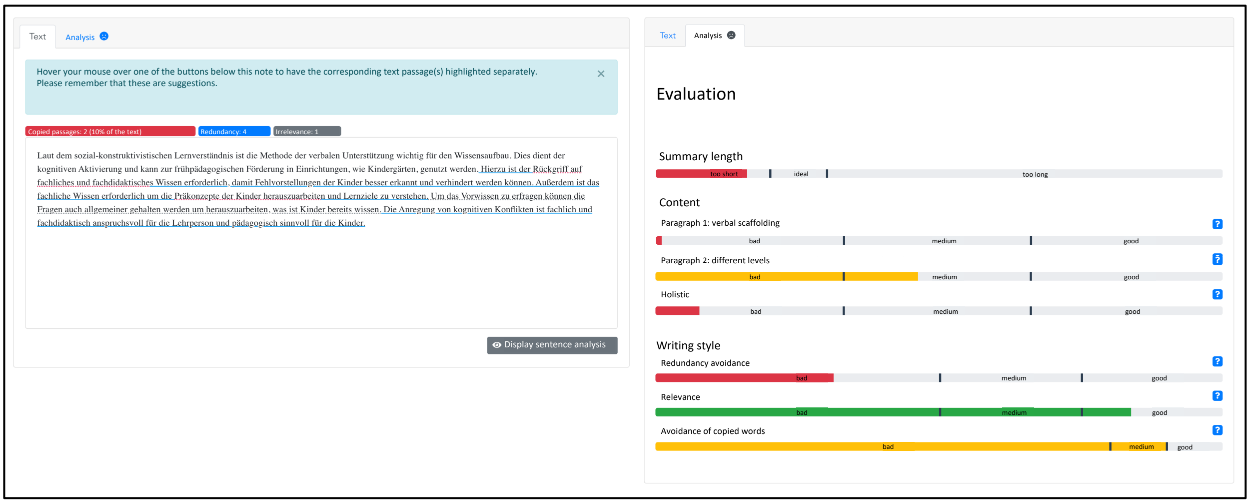


Figure 1

Example of Semantic and Score Feedback

Procedure. The study was conducted in an online tutorial during one semester with three experimental groups. The *prompt group* received a series of instructional prompts (Table 1) encouraging the use of summarizing strategies before being presented with the text editor to write down and revise the summary as many times as they wanted. The *feedback group* received semantic and score feedback after writing their first draft (Figure 1). The *combi group* received both treatments. The *feedback group* and the *combi group* were given information about how to decode the automated feedback and could ask for feedback and thus revise their first draft up to ten times.

Students were given two weeks to complete each assignment, which took approximately 90 minutes. At T0 and T5, the students submitted their summary but did not receive treatment. At T1 and T2, they received video lectures on the required aspects of text quality before summarizing.

Table 2
Treatment

T0	T1	T2	T3	T4	T5
Demographic data Language proficiency Pre-test	Video lecture: required aspects of text quality Treatment according to experimental condition	Video lecture: required aspects of text quality Treatment according to experimental condition	Treatment according to experimental condition	Treatment according to experimental condition	Post-test

3.3 Measures

Text quality was measured with a composed summary score consisting of five individual scores (content, plagiarism avoidance, redundancy avoidance, relevance, and length, i.e. Author, 2024). The range could theoretically span from 0 to 100. *Time on task* was measured in minutes from the time students were first able to read the text to the moment they submitted their final drafts. *Language proficiency* was assessed by self-reporting the grade received in the German class taken in the last year of high school. *Text difficulty* of the texts to be summarized was determined by a readability index for German texts (LIX; Lenhard & Lenhard, 2014). The LIX determines text difficulty by the sum of the average sentence length of a text and the percentage of long words (more than six letters).

3.4 Data Analysis

Analysis was conducted using R (R Core Team, 2022), version 4.3.2. Cases with less than 10% content or more than 90% text quality at T0 were removed from the analysis because it either indicates that the pre-test was not summarized properly, or students already possessed skills to write high quality summaries. To investigate the change in text quality as well as group differences between the experimental groups depending on text difficulty, language proficiency, and time on task, we specified multilevel models of change with person on level-2 and time of measurement on level-1. To do so, the data were organized in a long format to represent one time point per subject in each row. Thus, the unit of analysis was each measurement occasion for each student to model the change in the variables over time.

4 Results

Descriptive statistics for text quality and the correlational analysis are shown in Table 2 & 3. The main interest of this study was to analyse students' improvement in summarizing across six time points when receiving either automated formative feedback, instructional prompts, or a combination of both, according to the treatment condition.

Table 2
Descriptive Statistics of Text Quality, Time on Task, and Text Difficulty over Time

			T0		T1		T2		T3		T4		T5	
Text quality	N	Min/Max	M	sd	M	sd	M	sd	M	sd	M	sd	M	sd
Feedback	87	10/100	44.28	13.87	60.47	9.56	65.31	13.68	57.33	11.48	62.33	12.29	62.75	13.93
Prompt	75		50.73	14.55	57.32	12.33	62.51	13.06	53.17	14.56	56.13	14.53	58.79	17.59
Combi	92		46.10	17.05	62.10	11.39	68.30	13.31	60.88	11.42	65.87	7.58	64.08	14.47
Time on task														
Feedback	87	1/90	31.13	15.53	32.43	12.56	31.13	11.45	29.54	12.92	31.10	13.98	22.23	9.66
Prompt	75		35.86	12.23	40.85	15.24	-	-	37.12	13.83	37.79	16.33	21.38	9.53
Combi	92		35.01	14.81	35.57	14.10	28.79	12.86	32.67	15.81	29.07	16.37	23.66	10.96
Text difficulty			73.9		80.2		66.0		67.2		74.6		71.7	

Note. The maximum text quality in T0 was 90. Time on task was measured in minutes. In the prompt group, time on task in T2 was not measured due to a technical failure.

Table 3
Correlational Analysis of Text Quality, Time on Task, Language Proficiency, and Text Difficulty

Variable	1	2	3	4	5	6	7	8	9	10	11	12
1. Text quality t0												
2. Text quality t1	.19*											
3. Text quality t2	-.03	.29**										
4. Text quality t3	-.03	.23**	.06									
5. Text quality t4	.01	.16*	.13	.24**								
6. Text quality t5	.02	.13	.14	.21**	.00							
7. Time on task t0	.05	-.04	-.10	.05	.02	-.02						
8. Time on task t1	-.06	-.02	.04	.01	-.01	.00	.04					
9. Time on task t2	-.05	.02	-.04	.01	-.11	.06	-.03	.25**				
10. Time on task t3	-.08	.10	.08	.11	-.07	-.04	-.09	.26**	.10			
11. Time on task t4	.02	-.04	-.03	.02	-.08	.05	.07	.01	.14	.14		
12. Time on task t5	.03	.04	.04	.06	-.02	.18*	.00	.30**	.32**	.20*	-.03	
13. Language proficiency	-.13	-.08	.00	-.01	.01	.02	-.04	-.02	.05	.01	-.09	-.05

Note. * indicates $p < .05$. ** indicates $p < .01$.

First, we tested for a multilevel structure in the data by calculating the intraclass correlation. The results indicated that 8.9% of the variance in text quality over time could be attributed to individual differences, justifying a second level. Consequently, we employed multilevel modelling of change with measurement points nested in students to account for both interindividual and intraindividual change (Singer & Willett, 2003). To determine the optimal model, we tested various models incorporating different functions of time as fixed and random effects, utilizing the deviance statistic (Table 4). When time is treated as a fixed effect, the change in the dependent variable is set equal for all individuals, estimating differences only for individuals' intercepts (e.g., text quality at T0), but not for the rate of change. Conversely, when time is regarded as a random effect, both intercepts and the change in the dependent variable can vary between individuals.

Table 4
Model Comparisons

Model	Test of deviance
No time – Fixed time	$\chi^2 = 101.11$, $df = 1$, $p < .001$
Fixed time – fixed time^2	$\chi^2 = 81.60$, $df = 1$, p < .001
Fixed time^2 – Fixed time^3	$\chi^2 = 80.21$, $df = 1$, p < .001
Fixed time^3 – Random time	$\chi^2 = 15.91$, $df = 2$, p < .001
Random time – Random time^2	$\chi^2 = 11.95$, $df = 3$, p = .008
Random time^2 – Random time^3	—

Note. Time^2 = quadratic change in time. Time^3 = cubic change in time.

Tests of deviances showed that text quality followed a cubic change (time^3) and the effects of time differed between individuals. The model with a cubic slope as a random effect (time^3) did not converge, so the random time^2 model was chosen the best and more parsimonious model. Next, we included group, language proficiency, and text difficulty as level-1 fixed effects, and time on task as a level-1 random effect to analyse different rates of change between the feedback, prompt, and combi groups (Figure 3).

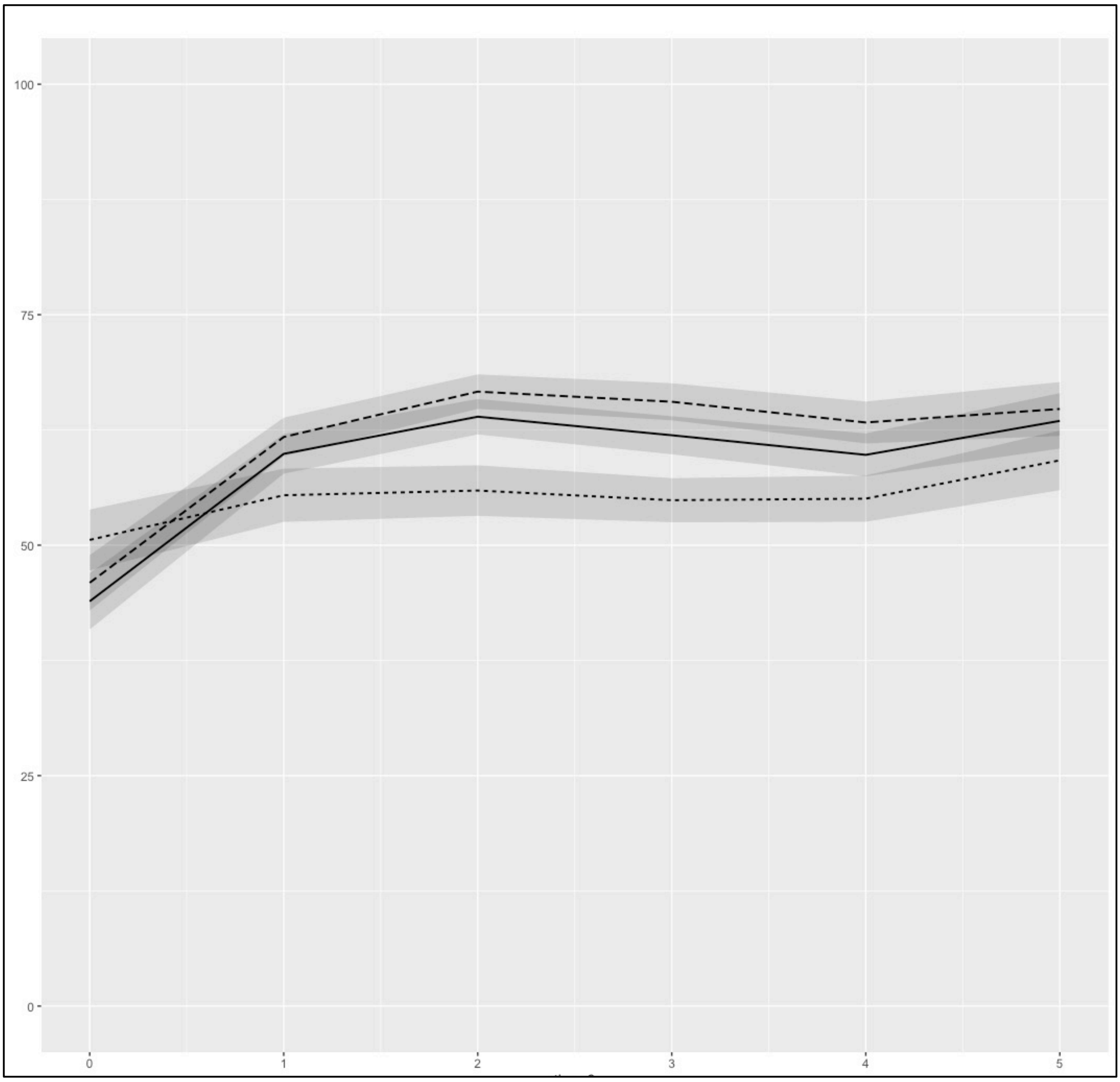


Figure 3
Predicted Values of Text Quality (y-axis) depending on Time (x-axis) for Feedback (solid), Combi (dashed), and Prompt (dotted) Groups

Text quality at T0 differed significantly between the prompt and the feedback groups ($\gamma_{\Delta\ prompt-feedback}$ = 6.66, p = .003) but not between the feedback and the combi groups ($\gamma_{\Delta\ combi-feedback}$ = 2.01, p = .350). When controlled for text difficulty, the feedback group's text quality increased substantially from T0 to T2 (γ_{time} = 23.96, p < .001), decreased from T2 to T4 ($\gamma_{time*time}$ = -8.95, p < .001), and then increased again slightly from T4 to T5 ($\gamma_{time*time*time}$ = 0.99, p < .001), thus following a cubic rate of change. The combi group exhibited a similar profile, with no significant differences observed in the slope ($\gamma_{\Delta time}$ = -1.08, p = .757; $\gamma_{\Delta time*time}$

$= 1.07$, $p = .525$; $\gamma_{\Delta time*time*time} = -0.18$, $p = .422$). The prompt group demonstrated a change profile comparable to that of the feedback and combi groups, albeit with only half the increase from T0 to T2 ($\gamma_{\Delta time} = -16.02$, $p < .000$), a slower decrease from T2 to T4 ($\gamma_{\Delta time*time} = 5.39$, $p = .005$), and a less pronounced increase from T4 to T5 ($\gamma_{\Delta time*time*time} = -0.53$, $p = .031$). Thus, the prompt group benefitted from the series of instructional prompts and wrote significantly better summaries compared to the pre-test. However, when compared to the feedback and combi groups, the prompt group exhibited a comparatively lower level of improvement in their summaries throughout the intervention. Neither language proficiency ($\gamma_{language\ proficiency} = -0.11$, $p = .842$) nor time on task ($\gamma_{time\ on\ task} = 0.00$, $p = .870$) had a significant impact on text quality.

Finally, we analysed the feedback and combination groups separately to assess the impact of providing additional prompts on students' engagement in feedback loops within the combi group. As a first step, we evaluated the relationship between feedback loops and text quality and found a positive association ($\gamma = 1.05$, $p = .001$) but no group differences ($\gamma_{\Delta\ feedback - combi} = 1.77$, $p = .420$). Next, we employed a mixed-effects model for feedback loops with person (level-2), group, language proficiency, text difficulty (all level-1 fixed effects), and time on task (level-1 random effect) as predictor variables (Table 5). The model showed that the number of feedback loops was not dependent on language proficiency, text difficulty, or time on task. However, the combi group consistently requested a slightly lower amount of feedback per session.

Table 5
Fixed Effects of Feedback Loops on Group, Language Proficiency, Text Difficulty, and Time on Task

Fixed effects	γ	*SE*	*p*
Δ Combi	-0.48	0.21	.022
Language proficiency	-0.04	0.12	.772
Text difficulty	-0.01	0.01	.368
Time on task	0.01	0.01	.027
R^2_{total}	.44		

5 Discussion

The present study was conducted to compare the interplay of different learning supports (instructional prompts and automated formative feedback) in promoting summarizing skills in an online tutorial for 254 undergraduate elementary education students. In detail, we analysed if prompts based on reciprocal teaching and automated formative feedback had a comparable impact on students' summarizing skills. Furthermore, we examined whether the combination of prompts and feedback holds additional advantages in promoting summarizing skills and if the combination has an impact on the requested amount of formative feedback. To this end, we implemented a pre-post-test design with 4 intervention points and three experimental groups. The *prompt group* was provided with instructional prompts for summarizing before using a text editor to write and revise summaries as often as they wanted to. The *feedback group* received semantic and score feedback during the writing process and could revise their first draft up to ten times. The *combi group* underwent both treatments.

Comparing the three experimental groups, we found that all groups improved text quality over the six sessions, yet to varying degrees. In the prompt condition, we examined the effectiveness of employing instructional prompts rooted in reciprocal teaching to support summarizing skills (Doolittle et al., 2006; Palinscar & Brown, 1984). These prompts aimed to activate prior knowledge, stimulate the assimilation of new information, and motivate the integration of text content into a broader context, thus fostering thoughtful reflection on texts and the construction of comprehensive mental models (Becker-Mrotzek et al., 2014; Hmelo-Silver & Eberbach, 2012; Perin et al., 2017). After having received these prompts, the students demonstrated improvement in their summaries throughout the intervention, maintaining elevated text quality in the post-test when not prompted, as compared to the pre-test. Consequently, our findings lead to the conclusion that the series of instructional prompts based on reciprocal teaching can enhance undergraduate university students' summarizing skills.

However, controlled for the performance in the pre-test, the learning gains of the prompt group were comparably low. Prompt group students were stimulated to intensively reflect on the content of the text and construct comprehensive mental models before starting to summarize. They could only evaluate their drafts through internal feedback relying on their individual reference standard. Consequently, these students may have lacked a detailed representation of an effective summary and may thus have faced challenges in revising sufficiently (i.e. Narciss, 2017). Moreover, they might have used the prompts merely as calls to action without deeper reflection (i.e. Engelmann et al., 2021). While the prompts explicitly addressed processes for knowledge construction, which should be reflected in a well-crafted summary (M. K. Kim & McCarthy, 2021; Schnotz, 2006), students may not have fully grasped their purpose in relation to summarizing, thus processing the task superficially, resulting in lower text quality compared to the groups that received feedback (de Silva & Graham, 2015).

The learning gains in the feedback and the combi group might be traced back to the automated feedback, that effectively pinpointed weaknesses in students' drafts and formatively confirmed their alignment with summary criteria. This might have led to a substantial increase in text quality from T0 to T1 which persisted throughout the intervention to the post-test. We conceive that the automated formative feedback prompted students to consistently evaluate their drafts against external standards, fostering an understanding of criteria that define a reference standard for good summaries.

The text quality of the combi group did not surpass that of the feedback group. This might imply a ceiling effect, suggesting that the use of automated formative feedback maximizes learning potential. The automated formative feedback given to the feedback and combi groups might have encouraged them to maintain a consistently high standard in their work (Author, 2024; Black & Wiliam, 2009). The immediate availability of a new feedback score after revision may have served as a direct incentive, encouraging careful draft revisions and the attainment of high text quality (Liu et al., 2017; Roscoe et al., 2015). This aligns with existing research indicating that seeking formative evaluation and interacting with automated formative feedback proves beneficial for students (Ali et al., 2018; Hyland & Zhang, 2018).

Yet, the introduction of supplementary prompts in the combi group resulted in a slight reduction in the number of feedback loops per session, indicating a potential substitution effect between the prompts and the feedback. In the combi group, where students were prompted to reflect on text content and construct a mental model, the use of summarizing strategies (e.g., selecting key aspects of a text) may have led to more sophisticated initial drafts. In contrast, students in the feedback group might not have intentionally employed summarizing strategies, resulting in comparatively less refined initial drafts. However, during the revision phase, feedback group students may have been able to elevate their drafts to a similar text quality as the combi group by engaging in more feedback loops. Conversely, combi group students may not have perceived the need for as many feedback loops as those in the feedback group to attain their desired feedback scores. They might have perceived the feedback more as a conclusive evaluation following the writing process rather than as formative assistance during the writing, as the feedback group may have interpreted it.

However, the level of engagement with feedback varies and is influenced by multiple factors, including task difficulty and the time invested in the activity (Manwaring et al., 2017; Zhang & Hyland, 2022). Yet, our study revealed that students' decision on the number of feedback loops did neither depend on the difficulty of the text to be summarized nor the time dedicated to the task.

5.1 Limitations

First, the four-session treatment period might have been insufficient to facilitate the adoption of effective summarizing strategies, thus the impact of our intervention could decline over time. Research has emphasized the significance of continuous training for acquiring summarizing strategies and skills (Friend, 2001; Kellogg, & Raulerson, 2007). Therefore, the treatment duration in this study may have been too brief, and the two-week processing time might have been too extended to fully realize the potential of the instructional prompts and automated formative feedback.

Second, the instructional prompts demonstrated a lower impact on the improvement of text quality than the automated feedback. This observation may, in part, be attributed to the wording of the prompts. We applied a content-oriented approach by adapting reciprocal teaching theory (Palinscar & Brown, 1984) to support the students' reading comprehension and mental modelling of the text. However, the inclusion of prompts such as asking for three examples to illustrate the text might have inadvertently misdirected the students. The algorithm which calculated the automated feedback prioritized the alignment of students' summaries with an expert summary of the original text rather than evaluating illustrated examples. Therefore, for future implementation, it is advisable to revise the prompts, placing a stronger emphasis prompting the alignment with the text quality standards of the tool.

Nonetheless, this study underscores that prompts designed from reciprocal teaching theory can promote summarizing, albeit to a lesser extent than automated formative feedback. In educational settings without access to a computer-based learning environment with such feedback, providing these prompts can serve as a viable alternative.

Third, this research applied a computer-based learning environment within an online tutorial on summarizing. Students were required to complete assignments within a two-week timeframe to meet course requirements. With students' freedom to choose when and where to work on the tasks, we faced limited control over ensuring strict adherence to instructions and proper engagement with the program. This constraint, however, enhances the ecological validity of the study, considering the prevalent use of online courses in universities. Nevertheless, we recognize that students may not have approached the prompts and feedback with the same level of seriousness as they might have in a more supervised setting.

5.2 Conclusion

The significance of possessing effective summarizing strategies for university students cannot be overstated, as summarizing enhances cognitive skills, communication abilities, and information processing. Given the prevalent absence of effective summarizing strategies among undergraduates, this research has shown how automated formative feedback and technology-enhanced learning environments can be deployed sensibly to provide more valuable learning opportunities for students. Furthermore, our study has effectively linked the enhancement of summarizing skills with instructional prompts and automated formative feedback and highlighted the complementary effect of instructional prompts and automated formative feedback within a computer-based learning environment. While the combined use of prompts and feedback may not directly lead to a greater improvement in students' summarizing proficiency compared to feedback alone, we understand that offering diverse access to promote summarizing strategies can create a more enriching experience in the computer-based learning environment. This approach may accommodate varied learning needs, fostering active engagement with text, and ultimately improving summarizing skills among students.

Acknowledgement

We express our gratitude to Prof. Dr. Wolfgang Lenhard for *conText,* his support in the development of *FALB*, and the enriching exchange. Additionally, we extend our thanks to Matthias Barde for the programming.

Funding

This project was part of the "Qualitaetsoffensive Lehrerbildung", a joint initiative of the Federal Government and the Laender that aims to improve the quality of teacher training [grant number 01JA2016]. The program is funded by the Federal Ministry of Education and Research, Germany. The authors are responsible for the content of this publication.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author.

References

Ahn, S. (2022). Developing summary writing abilities of Korean EFL university students through teaching summarizing skills. *English Teaching*, *77*(2), 25–43. https://doi.org/10.15858/engtea.77.2.202206.25

Ali, N., Ahmed, L., & Rose, S. (2018). Identifying predictors of students' perception of and engagement with assessment feedback. *Active Learning in Higher Education*, *19*(3), 239–251. https://doi.org/10.1177/1469787417735609

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* (second edition, pp. 316–329). Guildford Press.

Author. (2024).

Bannert, M. (2006). Effects of reflection prompts when learning with hypermedia. *Journal of Educational Computing Research*, *35*(4), 359–375. https://doi.org/10.2190/94V6-R58H-3367-G388

Becker-Mrotzek, M., Grabowski, J., jost, J., Knopp, M., & Linnemann, M. (2014). Adressatenorientierung und Kohärenzherstellung im Text: Zum Zusammenhang kognitiver und sprachlich realisierter Teilkomponenten von Schreibkompetenz. *Halbjahresschrift für die Didaktik der deutschen Sprache und Literatur*, *19*(37), 21–43.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, *21*(1), 5–31. https://doi.org/10.1007/s11092-008-9068-5

Brown, A., Palincsar, A., & Armbruster, B. B. (1984). Instructing comprehension-fostering Activities in interactive learning situations. *Learning and Comprehension of Text*, 255–286.

Butler, J. A., & Britt, M. A. (2010). Investigating instruction for improving revision of argumentative essays. *Written Communication*, *28*(1), 70–96. https://doi.org/10.1177/0741088310387891

Chew, C. S., Idris, N., Loh, E. F., Wu, W. V., Chua, Y. P., & Bimba, A. T. (2019). The effects of a theory-based summary writing tool on students' summary writing. *Journal of Computer Assisted Learning*, *35*(3), 435–449. https://doi.org/10.1111/jcal.12349 de Silva, R., & Graham, S. (2015). The effects of strategy instruction on writing strategy use for students of different proficiency levels. *System*, *53*, 47–59. https://doi.org/10.1016/j.system.2015.06.009

Deeva, G., Bogdanova, D., Serral, E., Snoeck, M., & De Weerdt, J. (2021). A review of automated feedback systems for learners: Classification framework, challenges, and opportunities. *Computers & Education*, *162*, 104094. https://doi.org/10.1016/j.compedu.2020.104094

Doolittle, P. E., Hicks, D., Triplett, C. F., Nichols, W. D., & Young, C. A. (2006). Reciprocal teaching for reading comprehension in higher education: A strategy for fostering the deeper understanding of texts. *International Journal of Teaching and Learning in Higher Education*, *17*(2), 106–118.

Duke, N. K., & Pearson, P. D. (2009). Effective practices for developing reading comprehension. *Journal of Education*, *189*(1–2), 107–122. https://doi.org/10.1177/0022057409189001-208

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

Engelmann, K., & Bannert, M. (2021). Analyzing temporal data for understanding the learning process induced by metacognitive prompts. *Learning and Instruction*, *72*, 101205. https://doi.org/10.1016/j.learninstruc.2019.05.002

Engelmann, K., Bannert, M., & Melzner, N. (2021). Do self-created metacognitive prompts promote short- and long-term effects in computer-based learning environments? *Research and Practice in Technology Enhanced Learning*, *16*(1), 1–21. https://doi.org/10.1186/s41039-021-00148-w

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

Friend, R. (2001). Effects of strategy instruction on summary writing of college students. *Contemporary Educational Psychology*, *26*(1), 3–24. https://doi.org/10.1006/ceps.1999.1022

Graham, S. (2006). Strategy instruction and the teaching of writing: A meta-analysis. In C. A. McArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of Writing Research* (pp. 187–207). The Guilford Press.

Graham, S. (2018). Instructional feedback in writing. In A. A. Lipnevich & J. K. Smith (Eds.), *The Cambridge Handbook of Instructional Feedback* (1st ed., pp. 145–168). Cambridge University Press. https://doi.org/10.1017/9781316832134.009

Green, J. M., & Holman, J. (2021). Cultivating the strategy of summarizing sequen-tialeExpository text: Scaffolds and supports for the intermediate grades. *Literacy Practice and Research*, *46*(1). https://doi.org/10.25148/lpr.009343

Hathorn, L. G., & Rawson, K. A. (2012). The roles of embedded monitoring requests and questions in improving mental models of computer-based scientific text. *Computers & Education*, *59*(3), 1021–1031. https://doi.org/10.1016/j.compedu.2012.04.014

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487

Hattie, J., Biggs, J., & Purdie, N. (1996). Effects of learning skills interventions on student learning: A meta-analysis. *Review of Educational Research*, *66*(2), 99–136. https://doi.org/10.3102/00346543066002099

Hidi, S., & Anderson, V. (1986). Producing written summaries: Task demands, cognitive operations, and implications for instruction. *Review of Educational Research*, *56*(4), 473–493. https://doi.org/10.3102/00346543056004473

Hill, M. (1991). Writing summaries promotes thinking and learning across the curriculum—But why are they so difficult to write? *Journal of Reading*, *34*(7), 536–539. http://www.jstor.org/stable/40014578

Hmelo-Silver, C. E., & Eberbach, C. (2012). Learning Theories and Problem-Based Learning. In S. Bridges, C. McGrath, & T. L. Whitehill (Eds.), *Problem-Based Learning in Clinical Education* (pp. 3–17). Springer Netherlands. https://doi.org/10.1007/978-94-007-2515-7_1

Hyland, K., & Zhang, Z. V. (2018). Student engagement with teacher and automated feedback on L2 writing. *Assessing Writing*, *36*, 90–102. https://doi.org/10.1016/j.asw.2018.02.004

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

Kim, K., Clarianay, R. B., & Kim, Y. (2019). Automatic representation of knowledge structure: Enhancing learning through knowledge structure reflection in an online course. *Educational Technology Research and Development*, *67*(1), 105–122. https://doi.org/10.1007/s11423-018-9626-6

Kim, M. K., & McCarthy, K. S. (2021). Improving summary writing through formative feedback in a technology-enhanced learning environment. *Journal of Computer Assisted Learning*, *37*(3), 684–704. https://doi.org/10.1111/jcal.12516

Kim, S.-A. (2001). Characteristics of EFL Rraders' summary writing: A study with Korean university students. *Foreign Language Annals*, *34*(6), 569–581. https://doi.org/10.1111/j.1944-9720.2001.tb02104.x

Kintsch, E. (1990). Macroprocesses and microprocesses in the development of summarization skill. *Cognition and Instruction*, *7*(3), 161–195. https://doi.org/10.1207/s1532690xci0703_1

Kintsch, E., Steinhart, D., Stahl, G., LSA Research Group, Matthews, C., & Lamb, R. (2000). Developing summarization skills through the use of LSA-based feedback. *Interactive Learning Environments*, *8*(2), 87–109. https://doi.org/10.1076/1049-4820(200008)8:2;1-B;FT087

Kirkland, M. R., & Saunders, M. A. P. (1991). Maximizing student performance in summary writing: Managing cognitive load. *TESOL Quarterly*, *25*(1), 105–121. https://doi.org/10.2307/3587030

Krause, U.-M., & Stark, R. (2010). Reflection in example- and problem-based learning: Effects of reflection prompts, feedback and cooperative learning. *Evaluation & Research in Education*, *23*(4), 255–272. https://doi.org/10.1080/09500790.2010.519024

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). Introduction to latent semantic analysis. *Discourse Processes*, *25*(2 & 3), 259–284. https://doi.org/10.1080/01638539809545028

Lee, Y. (2018). Effect of uninterrupted time-on-task on students' success in Massive Open Online Courses (MOOCs). *Computers in Human Behavior*, *86*, 174–180. https://doi.org/10.1016/j.chb.2018.04.043

Lehmann, T., Rott, B., & Schmidt-Borcherding, F. (2019). Promoting pre-service teachers' integration of professional knowledge: Effects of writing tasks and prompts on learning from multiple documents. *Instructional Science*, *47*(1), 99–126. https://doi.org/10.1007/s11251-018-9472-2

Lenhard, W., & Lenhard, A. (2014). *Berechnung des Lesbarkeitsindex LIX nach Björnson*. Unpublished. https://doi.org/10.13140/RG.2.1.1512.3447

Lenhard, W., Baier, H., Endlich, D., Lenhard, A., Schneider, W., & Hoffmann, J. (2012). Computerunterstützte Leseverständnisförderung: Die Effekte automatisch generierter Rückmeldungen. *Zeitschrift Für Pädagogische Psychologie*, *26*(2), 135–148. https://doi.org/10.1024/1010-0652/a000066

Lenhard, W., Baier, H., Lenhard, A., Hoffmann, J., & Schneider, W. (2013). *ConText: Förderung des Leseverständnisses durch das Arbeiten mit Texten: Manual*. Hogrefe.

Leopold, C., Brückner, A., & Dutke, S. (2019). Summarizing as a strategy for science text comprehension: Text-based versus content-based processing. *Discourse Processes*, *56*(8), 728–747. https://doi.org/10.1080/0163853X.2018.1563849

Lim, L., Bannert, M., van der Graaf, J., Singh, S., Fan, Y., Surendrannair, S., Rakovic, M., Molenaar, I., Moore, J., & Gašević, D. (2023). Effects of real-time analytics-based personalized scaffolds on students' self-regulated learning. *Computers in Human Behavior*, *139*, 107547. https://doi.org/10.1016/j.chb.2022.107547

Liu, M., Li, Y., Xu, W., & Liu, L. (2017). Automated essay feedback generation and its impact on revision. *IEEE Transactions on Learning Technologies*, *10*(4), 502–513. https://doi.org/10.1109/tlt.2016.2612659

Lüdecke, D. (2022). sjPlot: Data visualization for statistics in social science. R package version 2.8.12, https://CRAN.R-project.org/package=sjPlot

Manwaring, K. C., Larsen, R., Graham, C. R., Henrie, C. R., & Halverson, L. R. (2017). Investigating student engagement in blended learning settings using experience sampling and structural equation modeling. *The Internet and Higher Education*, *35*, 21–33. https://doi.org/10.1016/j.iheduc.2017.06.002

McAnulty, S. J. (1981). Paraphrase, summary, precis: Advantages, definitions, models. *Teaching English in the Two-Year College*, *8*(1), 4751.

Narciss, S. (2017). Conditions and effects of feedback viewed through the lens of the interactive tutoring feedback model. In D. Carless, S. M. Bridges, C. K. Y. Chan, & R. Glofcheski (Eds.), *Scaling up Assessment for Learning in Higher Education* (Vol. 5, pp. 173–189). Springer Singapore.

Nixon, S., Brooman, S., Murphy, B., & Fearon, D. (2016). Clarity, consistency and communication: Using enhanced dialogue to create a course-based feedback strategy. *Assessment & Evaluation in Higher Education*, *42*(5), 812–822. https://doi.org/10.1080/02602938.2016.1195333

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

Perin, D. (2019). Reading, writing, and self-efficacy of low-skilled postsecondary students. In D. Perin (Ed.), *The Wiley Handbook of Adult Literacy* (1st ed., pp. 237–260). Wiley. https://doi.org/10.1002/9781119261407.ch11

Perin, D., Lauterbach, M., Raufman, J., & Kalamkarian, H. S. (2017). Text-based writing of low-skilled postsecondary students: Relation to comprehension, self-efficacy, and teacher judgments. *Reading and Writing*, *30*(4), 887–915. https://doi.org/10.1007/s11145-016-9706-0

Posit Team (2023). RStudio: Integrated development environment for R. Posit Software, PBC, Boston, MA. http://www.posit.co/

Proske, A., Narciss, S., & McNamara, D. S. (2012). Computer-based scaffolding to facilitate students' development of expertise in academic writing. *Journal of Research in Reading*, *35*(2), 136–152. https://doi.org/10.1111/j.1467-9817.2010.01450.x

R Core Team (203). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Roscoe, R. D., Snow, E. L., Allen, L. K., & McNamara, D. S. (2015). Automated detection of essay revising patterns: Applications for intelligent feedback in a writing tutor. *Technology, Instruction, Cognition and Learning*, *10*, 59–79.

Schnotz, W. (2006). Was geschieht im Kopf des Lesers? Mentale Konstruktionsprozesse beim Textverstehen aus der Sicht der Psychologie und der kognitiven Linguistik. *Text-Verstehen. Grammatik und darüber hinaus*, 222–238.

Schunk, D. H., & Rice, J. M. (1991). Learning Goals and Progress Feedback during Reading Comprehension Instruction. *Journal of Reading Behavior*, *23*(3), 351–364. https://doi.org/10.1080/10862969109547746

Seifried, E., Lenhard, W., Baier, H., & Spinath, B. (2012). On the reliability and validity of human and LSA-based evaluations of complex student-authored texts. *Journal of Educational Computing Research*, *47*(1), 67–92. https://doi.org/10.2190/EC.47.1.d

Shanahan, T. (2019). Reading—Writing connections. In *Best Practices in Writing Instruction.* Guilford Press.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153–189. https://doi.org/10.3102/0034654307313795

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford University Press.

Stanley, D. (2021). apaTables: Create American psychological association (APA) style tables. R package version 2.0.8, https://CRAN.R-project.org/package=apaTables

Stevens, E. A., Park, S., & Vaughn, S. (2019). A review of summarizing and main idea interventions for struggling readers in grades 3 through 12: 1978–2016. *Remedial and Special Education*, *40*(3), 131–149. https://doi.org/10.1177/0741932517749940

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 & Education*, *131*, 33–48. https://doi.org/10.1016/j.compedu.2018.12.005

Sung, Y.-T., Liao, C.-N., Chang, T.-H., Chen, C.-L., & Chang, K.-E. (2016). The effect of online summary assessment and feedback system on the summary writing on 6th graders: The LSA-based technique. *Computers & Education*, *95*, 1–18. https://doi.org/10.1016/j.compedu.2015.12.003

Teng, M. F. (2022). Effects of individual and group metacognitive prompts on tertiary-level students' metacognitive awareness and writing outcomes. *The Asia-Pacific Education Researcher*, *31*(5), 601–612. https://doi.org/10.1007/s40299-021-00611-8

Tolosa, C., East, M., & Villers, H. (2015). Motivating Twenty-First-Century Learners: The Impact of an Online Reciprocal Peer-Tutoring Initiative for Foreign Language Learning. In C. Koh (Ed.), *Motivation, Leadership and Curriculum design* (pp. 137–149). Springer Singapore. https://doi.org/10.1007/978-981-287-230-2_11

Topping, K. J., Dehkinet, R., Blanch, S., Corcelles, M., & Duran, D. (2013). Paradoxical effects of feedback in international online reciprocal peer tutoring. *Computers & Education*, *61*, 225–231. https://doi.org/10.1016/j.compedu.2012.10.002

van den Boom, G., Paas, F. G. W. C., & van Merriënboer, J. J. G. (2007). Effects of elicited reflections combined with tutor or peer feedback on self-regulated learning and learning outcomes. *Learning and Instruction*, *17*(5), 532–548. https://doi.org/10.1016/j.learninstruc.2007.09.003

van den Boom, G., Paas, F. G. W. C., van Merriënboer, J. J. G., & van Gog, T. (2004). Reflection prompts and tutor feedback in a web-based learning environment: Effects on students' self-regulated learning competence. *Computers in Human Behavior*, *20*(4), 551–567. https://doi.org/10.1016/j.chb.2003.10.001

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. Academic Press.

Wade-Stein, D., & Kintsch, E. (2004). Summary Street: Interactive computer support for writing. *Cognition and Instruction*, *22*(3), 333–362. https://doi.org/10.1207/s1532690xci2203_3

Westby, C., Culatta, B., Lawrence, B., & Hall-Kenyon, K. (2010). Summarizing expository texts. *Topics in Language Disorders*, *30*(4), 275–287. https://doi.org/10.1097/TLD.0b013e3181ff5a88

Zellermayer, M., Salomon, G., Globerson, T., & Givon, H. (1991). Enhancing writing-related metacognitions through a computerized writing partner. *American Educational Research Journal*, *28*(2), 373–391.

Zhang, Z., & Hyland, K. (2022). Fostering student engagement with feedback: An integrated approach. *Assessing Writing*, *51*, 100586. https://doi.org/10.1016/j.asw.2021.100586

Zheng, L. (2016). The effectiveness of self-regulated learning scaffolds on academic performance in computer-based learning environments: A meta-analysis. *Asia Pacific Education Review*, *17*(2), 187–202. https://doi.org/10.1007/s12564-016-9426-9

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

Zimmerman, B. J., & Kitsantas, A. (2007). A writer's discipline: The development of self-regulatory skill. In G. Rijlaarsdam, P. Boscolo, & S. Hidi (Eds.), *Studies in Writing* (Vol. 19, pp. 51–69). Elsevier.

Appendix

Appendix A: Multilevel Models of Change

Table A1
Multilevel Model of Change in Text Quality

Fixed effects	*Estimates*	*SE*	*p*
Intercept (FB)	29.49	6.48	.000
Time (FB)	23.96	2.50	.000
$Time^2$ (FB)	-8.95	1.21	.000
$Time^3$ (FB)	0.99	0.16	.000
Intercept ΔPT	6.66	2.27	.003
Time ΔPT	-16.02	4.03	.000
$Time^2$ ΔPT	5.39	1.90	.005
$Time^3$ ΔPT	-0.53	0.24	.031
Intercept ΔCombi	2.01	2.15	.350
Time ΔCombi	-1.08	3.49	.757
$Time^2$ ΔCombi	1.07	1.68	.525
$Time^3$ ΔCombi	-0.18	0.22	.422
Time on task	0.00	0.03	.870
Language proficiency	-0.11	0.54	.842
Text difficulty	0.20	0.08	.011
Random Effects	*Var*	*SD*	
Person	70.11	8.01	
Time	19.62	4.43	
$Time^2$	0.94	0.97	
Level-1 residual	139.89	11.83	
$R^2_{marginal/conditional}$	0.184 / 0.367		

Table A2
Multilevel Model of Change in Feedback Loops

Fixed effects	*Estimates*	*SE*	*p*
Intercept (FB)	2.94	0.80	.000
ΔCombi	-0.48	0.21	.022
Time on task	0.01	0.01	.027
Language proficiency	-0.04	0.12	.772
Text difficulty	-0.01	0.01	.368
Random Effects	*Var*	*SD*	
Person	1.44	1.20	
Level-1 residual	1.96	1.40	
$R^2_{marginal/conditional}$	0.022 / 0.437		

Note: The **marginal R^2** describes the proportion of variance explained by the fixed factor(s) alone. The **conditional R^2** describes the proportion of variance explained by both the fixed and random factors.

Curriculum Vitae

VERONIKA BARKELA

CAREER

Research associate

Since 2024/01	**RPTU Kaiserslautern-Landau** | Landau, Germany (former University of Koblenz-Landau) *Institute of Children and Youth Education* Project ComE, fostering computational thinking in elementary education students. Development of an automated feedback system, scaffolding, and learning materials, investigation of students' learning processes, motivation, self-regulation, and feedback behavior.
Since 2016/08	Project FALB, fostering summarizing skills with automated formative feedback. Development of an automated feedback system and prompts, investigation of students' learning processes, behavior, and motivation.
Since 2023/04	Conception, planning, and realization of science seminars for elementary education students, supervising master thesis.
2020/03 – 2023/12	Project ForALSA, funded by the federal government and the laender of Germany. Development of test instruments, investigation of students' summarizing skills.
2016/10 – 2020/02	Conception, planning, and realization of science seminars for elementary education students, supervising master thesis.
Coordinator 2013/09 – 2015/07	**University of Koblenz-Landau** | Landau, Germany *Career Service* Conception, organization, and realization of trainings for university students and staff

EDUCATION

Doctoral studies Since 2016/10	**RPTU Kaiserslautern-Landau** | Landau, Germany (former University of Koblenz-Landau) As a social scientist in educational studies
Degree in Social Sciences 2003/10 – 2013/12	**University of Koblenz-Landau** | Landau, Germany Title of the master thesis: “Intercultural inclusion in public institutions”

TEACHING EXPERIENCE

LECTURER Since 2024/10	**RPTU Kaiserslautern-Landau** | Landau, Germany Seminar on democracy education for students in the master's program in elementary education. Topics: pluralism, multiculturality, racism, antisemitism, antigypsyism
Since 2023/04 and 2016/10-2020/02	**RPTU Kaiserslautern-Landau** | Landau, Germany (former University of Koblenz-Landau) 2-3 seminars per semester on inquiry-based learning for students in the master's program in elementary education. Topics: computational thinking, magnetism, electricity, building bridges
Thesis advisor Since 2016/10	**RPTU Kaiserslautern-Landau** | Landau, Germany (former University of Koblenz-Landau) 3–6 master's theses per semester Topics: conceptual change, scaffolding, self-regulation, motivation, and self-efficacy.

FUNDING & AWARDS

08/2021	**Best paper award** EARLI-JURE 2021 (1. Platz): Barkela, V. & Leuchter, M. (2021). *Promoting Student Teachers' Summarizing Skills through Automated Feedback and Self-Assessment.*
07/2020	**Teaching award** Hochschuldidaktische Arbeitsstelle of the University of Koblenz-Landau, Germany
07/2018	**Teaching award** Hochschuldidaktische Arbeitsstelle of the University of Koblenz-Landau, 500€
10/2017 – 08/2018	**University of Koblenz-Landau** | Landau, Germany Mentee in the mentoring program ment^2 für PhD students
2017	**ERASMUS+** Course on scientific English, Malta, 900€

REVIEWING ACTIVITIES

Since 2024	Computers & Education (CiteScore 27.1) Contemporary Educational Psychology (CiteScore 16.5) European conference on information systems (ECIS)

COMMITTEE WORK

Since 2024	**RPTU Kaiserslautern-Landau** | Landau, Germany Representative of doctoral candidates on the HRS4R steering committee
03/2022 – 09/2024	**RPTU Kaiserslautern-Landau** | Landau, Germany Doctoral representation (Speaker of the Department)

KNOWLEDGE & SKILLS

Languages:	German (native language) English (business fluent) French, Norwegian (basic)
Computer skills:	MS-Office, SPSS, R (advanced)

PUBLICATION LIST

- **Barkela, V.**, & Leuchter, M. (2024). The interplay of instructional prompts and automated formative feedback. *Manuscript submitted for publication.*
- Stiel-Dämmer, S., **Barkela, V.**, & Leuchter, M. (2024). Die Förderung des Erkennens von kognitiver Aktivierung durch den Einsatz von Vergleichen. *Manuscript submitted for publication.*
- Weber, A. M., **Barkela, V.**, Han, A., Mühling, A., Greiff, S., & Leuchter, M. (2024). Dynamics of Programming Anxiety and Programming Enjoyment in Preservice Teachers: The Role of Academic Self-Concept and Gender. *Manuscript submitted for publication.*
- **Barkela, V.**, Han, A., & Weber, A.M. (2024). Do student teachers experience self-worth threats in computational thinking? *Computers in Human Behavior Reports*, 100463. https://doi.org/10.1016/j.chbr.2024.100463
- **Barkela, V.,** & Leuchter, M. (2024). Effectiveness of automated formative feedback in an online tutorial for promoting summarizing. *Journal of Educational Technology Development and Exchange, 17*(1), 62-90. doi.org/10.18785/jetde.1701.04.

- **Barkela, V.**, Schmitt, L., & Leuchter, M. (2023). The impact of cognitive and motivational resources on engagement with automated formative feedback. *Contemporary Educational Psychology*, 102234. https://doi.org/10.1016/j.cedpsych.2023.102234
- Weber, A., Bastian, M., **Barkela, V.**, Mühling, A., & Leuchter, M. (2022). Fostering preservice teachers' expectancies and values towards computational thinking. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.987761
- Weber, A. M., **Barkela, V.,** Stiel-Dämmer, S., & Leuchter, M. (2021). Der Zusammenhang emotionaler Kosten bei Grundschullehramtsstudierenden mit ihrer informatischen Problemlösekompetenz. *Empirische Pädagogik, 35*(1), 93–111. (Open Access)
- **Barkela, V.,** & Leuchter, M. (2019). Förderung der Überarbeitung von Textzusammenfassungen sachunterrichtlicher Fachtexte durch Self-Assessment und Peer Feedback mit Prompts. *GDSU-Journal, 9*, 139–153.

Conference contributions (selection)

- **Barkela, V.,** & Leuchter, M. (2024). Student Agency in E-learning with automated Formative Feedback. *Poster presented at the annual meeting of SIG7 of the European Association for Research on Learning and Instruction,* Tübingen, GER: 21.08. - 23.08.2024.
- **Barkela, V.**, Han, A., & Weber, A.M. (2024). Do Student Teachers Experience Self-worth Threats in Computational Thinking? *Paper presented at JURE 2024*, Sevilla, SP: 24.06. - 28.06.2024.
- **Barkela, V.,** Schmitt, L., & Leuchter, M. (2023). Reciprocal Relationships of Motivation, Mental Effort, and Behavioral Feedback Engagement. *Paper presented at the biannual meeting of the European Association for Research on Learning and Instruction*, Thessaloniki, GR: 22.08. - 26.08.2023.

- **Barkela, V.,** & Leuchter, M. (2023). FALB - automatisiertes Feedback zur Förderung der Zusammenfassungskompetenz. *Präsentation auf der Veranstaltung "KI in der Lehrer:innenbildung. Disruptive Kraft oder Zukunftskompetenz"*, Mainz, GER: 14.06.2023.
- **Barkela, V.,** & Leuchter, M. (2022). Der Einfluss von motivationalen und kognitiven Aspekten auf die individuelle Nutzung formativen, automatisierten Feedbacks. *Präsentation auf der DiLeBi-Tagung*, Mainz, Trier, Koblenz, Landau, GER: 10. – 14.10.2022.
- **Barkela, V.,** & Leuchter, M. (2021). Promoting Student Teachers' Summarizing Skills through Automated Feedback and Self-Assessment. *Paper presented at the biannual meeting of the Junior Researchers of the European Association for Research on Learning and Instruction*, Gothenburg, SWE: 18.08. - 20.08.2021.
- **Barkela, V.,** & Leuchter, M (2021). Zusammenfassen von fachwissenschaftlichen Texten mit computerbasiertem formativem Assessment fördern. *Poster auf dem Programmworkshop im Rahmen der "Qualitätsoffensive Lehrerbildung"*. Frankfurt, GER: 24.06. - 25.06.2021.
- **Barkela, V.,** & Leuchter, M (2021). Fostering summarizing with computer-based formative assessment in science teacher education. *Poster auf der Schwerpunkttagung der Gesellschaft für Didaktik der Chemie und Physik zu maschinellem Lernen und computerbasierten Textanalysen*. 06.05. - 07.05.2021.

Landau, den 13.12.2024

Veronika Barkela

Declaration of Originality

I confirm that I wrote this dissertation independently, without outside help, and only used the resources listed. While working on this dissertation, I used DeepL Write and ChatGPT to improve the language and readability. The AI tools did not write paragraphs or generate content but were used to refine the text. I reviewed and edited after using these tools and take full responsibility for the content of this dissertation. This dissertation has not been submitted for grading at this or any other university before.

Landau, 13.12.2024

Veronika Barkela