Bridging Domains: Examining the Effects of Relevance Instructions and Guiding Questions on Pre-Service Teachers' First- and Second-Order Knowledge Integration

Thomas Lehmann

University of Bremen, Faculty of Pedagogy and Educational Sciences, Research Unit: Learning, Instruction, and Educational Psychology, Universitäts-Boulevard 11/13, 28359 Bremen, Germany

Correspondence concerning this article should be addressed to Thomas Lehmann, University of Bremen, Faculty 12: Pedagogy and Educational Sciences, Universitäts-Boulevard 11/13, 28359 Bremen, Germany, Phone: +49 421 218 69130, E-Mail: tlehmann@uni-bremen.de

Abstract

The conceptualization of pre-service teachers' knowledge integration typically involves the distinction of two types: first-order knowledge integration, which includes merging domain-specific knowledge entities into a common knowledge base, and second-order knowledge integration, which refers to the integrated (simultaneous) application of knowledge from diverse domains. This study investigates the effect of instructional prompts in the form of (a) relevance instructions and (b) guiding questions on promoting pre-service teachers' first-order knowledge integration in a reading- and writing-based learning setting with three domain-specific study texts: one text each referring to content knowledge (CK), general pedagogical knowledge (PK), or pedagogical-content knowledge (PCK). Furthermore, the study explores whether pre-service teachers' second-order knowledge integration depends on the degree to which they engaged in first-order knowledge integration when reading and writing about different domain-specific learning contents. The study applied a three parallel group experimental design. An analysis of essays written by $N = 83$ German language pre-service teachers indicated positive effects of both prompts on first-order knowledge integration. Moreover, a mediation analysis showed that pre-service teachers' second-order knowledge integration is mediated by their first-order knowledge integration. The results are discussed and integrated into the existing body of research, practical implications are presented, and limitations of the study are explained.

Keywords: Knowledge integration; Prompts; Guiding questions; Relevance instructions; Multiple document comprehension; Pre-service teacher education.

Introduction

The success of learning in higher education is heavily contingent on proficiency in studying through a multitude of reading materials and writing. Therefore, students are usually accustomed to engaging in learning activities that necessitate the retrieval, comparison, and amalgamation of information from various written documents (Mateos et al., 2018) and to composing their own texts for the sake of deep comprehension (Tynjälä et al., 2001). Nevertheless, research on multiple document comprehension indicates that students frequently encounter difficulties in executing the cognitive processes necessary for effective knowledge integration across different documents (Bigot & Rouet, 2007), especially when these documents originate from diverse domains and provide complementary information (e.g., Lehmann et al., 2019; Wäschle et al., 2015). This presents a significant challenge in teacher education because pre-service teachers must acquire knowledge of diverse domains, including subject-matter knowledge of the subject(s) they will teach as in-service teachers (content knowledge; CK), generic pedagogical knowledge (PK), and knowledge about subject didactics (pedagogical content knowledge; PCK) (Baumert & Kunter, 2006; Shulman, 1986). However, research which adopted the expert–novice paradigm shows that expert teachers differ from novices not only in providing vast knowledge bases in each of these domains; they also have the different bodies of knowledge elaborated and organized into a well-integrated structure (Berliner, 2001; Bromme, 2014; König, 2010; Krauss, 2011; Lachner et al., 2016; Livingston & Borko, 1990). Pre-service teachers, on the other hand, appear to have their CK, PK, and PCK compartmentalized into separate memory parts (Harr et al., 2014; Renkl, et al., 1996). This results in difficulties in adopting multiple domain-specific perspectives on solving tasks or problems (Weinert et al., 1990), for example in lesson planning (Janssen & Lazonder, 2016) and the design of instructional assignments (Wäschle et al., 2015). Yet, the structure of teacher training programs in various countries is usually not conducive to pre-service teachers' knowledge integration, as the teaching of the different knowledge bodies is spread

across different courses or even different institutions (Darling-Hammond, 2006; Hudson & Zgaga, 2017). It is therefore essential to ensure curricular means of fostering knowledge integration in pre-service teacher education, such as a systematic coordination of the learning content of CK-, PK-, and PCK-specific courses. However, even though such changes are now slowly being realized (e.g., Lilliedahl et al., 2020; Schellenbach-Zell & Neuhaus, 2022), they come with certain obstacles that make the implementation difficult. For example, such changes require extensive organizational effort and also an increase in faculty members' understanding of the necessity of integrated teaching of distinct knowledge bodies (Zeeb et al., 2019). Therefore, researchers previously merged two research fields, namely teacher professionalization and multiple document comprehension, and emphasized rather easy-to-implement instructional scaffolds such as *prompts* to promote pre-service teachers' knowledge integration within self-regulated learning scenarios. These scenarios involve engagement with various domain-specific texts encompassing CK, PK, and PCK, as well as the practice of writing to facilitate learning (Tynjälä et al., 2001; see also Lehmann, 2020a). This approach aligns with the methodology employed in the present study.

Prompts are externally provided "performance aids" (Bannert, 2009, p. 139) that direct learners' attention to specific aspects of the learning content and/or stimulate certain activities, such as elaborative thinking, metacognitive planning, monitoring and control, goal focusing, and self-motivation, to facilitate task completion and learning. Prompts can take on different forms, such as questions, explicit statements, sentence-starters, execution instructions, pictures, graphics, and multimodal forms (Bannert, 2009; Lehmann et al., 2014). Recently, *guiding questions* (also called focus questions) have been successfully used to prompt knowledge integration in the context of reading- and writing-based learning environments for pre-service teachers (e.g., Lehmann et al., 2019, 2020; Lehmann, 2020a; Wäschle et al., 2015). *Relevance instructions* are another type of "externally provided prompts" (McCrudden & Schraw, 2010, p. 97). However, unlike other prompts, relevance

instructions are specifically designed to make a certain goal relevant to learners prior to their actual learning and performance phase (McCrudden & Schraw, 2007, 2011a). Hence, relevance instructions differ from guiding questions in that they are not intended to elicit specific responses or stimulate particular processing modes. For example, a relevance instruction may explain the importance of integrating knowledge from different domains before learners engage with multiple sources. By this means, a relevance instruction directly aids pre-service teachers' task model formation (Britt et al., 2017; McCrudden & Schraw, 2007, 2010), allowing them to subsequently allocate their resources and use appropriate strategies for the integration of different knowledge bodies (Zeeb et al., 2020). Guiding questions, on the other hand, are considered sense-making *directed prompts*, as they directly provoke learners to engage in certain processing modes by answering the questions (Davis, 2003).

Given that both relevance instructions and guiding questions improve pre-service teachers' integration of CK, PK, and PCK in learning, these instructional means should, from a theoretical perspective, also promote the simultaneous, integrated use of the different knowledge bodies in application contexts (Lehmann, 2020b). However, empirical studies on the relation between knowledge integration in learning and in application settings are scarce and findings vary. While the study of Zeeb et al. (2019) provides evidence that a relevance instruction increased the integrated use of two different knowledge domains in scenario-based tasks, it remains open whether the effect is limited to the domains considered (PK and PCK) and/or to specifics of the computer-based learning environment of this study, which incorporated two domain-specific video-lectures. By contrast, guiding questions were found to enhance integrative learning processes that facilitate the interrelation and merging of domain-specific knowledge entities into a unified structure (e.g., Lehmann et al., 2019, 2020; Wäschle et al., 2015). But here, findings on the integrated use of domain-specific knowledge are rare. Therefore, the present study investigated the immediate effects of relevance

instructions and guiding questions on knowledge integration as a form of (a) first-order knowledge integration, which includes merging domain-specific knowledge entities into a common knowledge base, and (b) second-order knowledge integration, which refers to the integrated simultaneous application of CK, PK, and PCK. More specifically, the study examined whether relevance instructions and guiding questions promote pre-service teachers' first-order knowledge integration in a reading- and writing-based learning setting with multiple domain-specific study texts and whether the potential effects of these prompts on the integrated use of CK, PK, and PCK in an application test (second-order knowledge integration) are mediated by pre-service teachers' integrative learning performance (first-order knowledge integration).

Knowledge integration in pre-service teachers

Interest in the construct of *knowledge integration* (KI) in pre-service teachers has grown considerably over the last two decades. Still, KI is often not consistently conceptualized. Some researchers consider KI to be "a process of integrating what is known into action" (Gottein, 2020, p. 231). Other researchers perceive KI as a process of interrelating, combining, and merging originally unconnected pieces and structures of knowledge across different topics and domains to build an integrated knowledge base (e.g., Lee & Turner, 2017; Schneider, 2012).

It can be argued that the two notions are just different sides of the same coin. While one side refers to the cognitive-constructive learning processes of building and structuring an integrated knowledge base within one's memory (i.e., *first-order* KI), the other side denotes the integrated use of domain-specific knowledge in due consideration of how particular knowledge entities draw on or interact with each other as regards teaching (i.e., *second-order* KI) (Lehmann, 2020b). This perspective exhibits a strong connection with the concept of transfer, which is commonly understood as the skillful application of prior learning in (rather) unfamiliar situations or tasks (e.g., Gick & Holyoak, 1987; Hajian, 2019). If, for example, a

pre-service teacher intends to introduce a new topic using visualizations in class, they will profit from having learned which representations are not only correct from a CK standpoint but also suitable considering learners' cognitive processing (PK) and topic-specific preconceptions (PCK) (Graichen et al., 2019). Assuming that the learning involved reading multiple domain-specific texts, which is probably one of the most common forms of learning in the academic phase of teacher education, the decision of how to introduce the new topic using visualizations represents a far transfer (because the learning and the transfer situation are apart and completely different; Perkins & Salomon, 1992).

In addition to these correspondences with theories on learning transfer, Lehmann's (2020b) conceptualization of KI matches models of teachers' professional competence (e.g., Blömeke et al., 2015). First-order KI shapes a (prospective) teacher's cognition as a disposition for their professional competence in that it affects the degree to which the knowledge is represented in an integrated manner in their memory. The underlying cognitive processes mainly involve elaboration and critical thinking (Lee & Turner, 2017; Lehmann, 2021), but also an integrative application of organizing the learning content of various, potentially domain-specific, sources (Wäschle et al., 2015). Second-order KI can be associated with the connecting processes (i.e., perception, interpretation, and decision-making) between an individual's cognition as a dispositional trait and their performance in particular situations. Each of the skills that mediates between disposition and performance demands the concurrent consideration and application of CK, PK, and PCK. Following the same line of reasoning, Zeeb and colleagues state that "integrated knowledge structures are therefore an important prerequisite for the practical application in the classroom" (Zeeb et al., 2020, p. 203). This view applies not only to classroom teaching but also to lesson planning and the design of learning tasks and material. Yet, pre-service teachers typically struggle in implementing effective integration strategies such as the integrative elaboration of domain-specific information, especially when they are not stimulated to engage in such endeavors

(e.g., Lehmann et al., 2019). It is thus not surprising that they often lack an integrated application and rather rely on a single knowledge domain in practice. This is backed up by evidence from qualitative (Seel, 1997) and mixed-methods studies (Wäschle et al., 2015) that pre-service teachers predominantly rely on CK for lesson planning and the design of learning tasks.

Aside from rather "high-threshold" curricular changes that foster the merging of domain-specific into integrated university courses and practical elements (e.g., Barzel et al., 2016; Lilliedahl et al., 2020; Schellenbach-Zell & Neuhaus, 2022), there are two common approaches to promoting KI in pre-service teachers. The first approach refers to providing learning environments and material that incorporate various domain-specific learning contents to be integrated by pre-service teachers already in an integrated manner. That is, instructional designers and/or lecturers try to make KI happen, at least in parts, prior to pre-service teachers' learning, which subsequently reduces the complexity of the integration task (see Renkl, 2014). Various studies suggest that these approaches have positive effects on pre-service teachers' first- and second-order KI (e.g., Harr et al., 2014; Janssen & Lazonder, 2016). However, designing and implementing such integrated learning environments require a lot of effort (see Hudson & Zgaga, 2017; Lilliedahl et al., 2020).

The second, rather "low-threshold" approach refers to posing tasks and providing task-supplemental instructional scaffolds, such as prompts. Both of these means, tasks and prompts, can be specifically designed to promote pre-service teachers' integrative knowledge building and structuring (e.g., Lehmann et al., 2019). Tasks challenge learners to work out a measurable outcome (i.e., a solution to the task). The process of completing the task provides opportunities to learn something (e.g., acquire knowledge, gain deeper understanding, develop certain skills). Prompts, on the other hand, differ from tasks in that they are "non-standalone" means. Rather, they assist learners in successfully completing a task at hand and/or attaining a specific goal. That is, prompts enable learners to engage in complex tasks

that would be overstraining without any instructional support (Bannert, 2009; Rosenshine & Meister, 1992). This is a critical issue in learning environments that involve multiple domain-specific documents, as learners may formulate an adequate goal for KI but lack a comprehensive understanding of the actions, procedures, and strategies necessary to achieve it (Linderholm et al., 2014). This knowledge gap can impede learners' ability to integrate relevant information and construct an integrated understanding of the—potentially domain-specific—learning content. Therefore, research on (multiple) document comprehension aimed at identifying scaffolds that support learners' integrative processing, either directly by posing certain questions (e.g., Moreno et al., 2020; Rouet et al., 2001; Smith et al., 2010) or indirectly by enhancing learners' task model with task-supplemental relevance instructions (e.g., McCrudden et al., 2007, 2011b).

For example, Rouet et al. (2001) found that low-level guiding questions encouraged a "locate-and-memorize" strategy where students only focused on text segments that directly corresponded to the question, while higher-level guiding questions prompted a "review-and-integrate" strategy where students focused more broadly on sections of the text that contained information relevant to reflecting on the question.

Relevance instructions address the ideas of purposeful reading (Britt et al., 2017) and goal-focusing (McCrudden & Schraw, 2007) contained in the REading as problem SOLVing (RESOLV) framework, which builds on previous models of purposeful reading (e.g., Rouet and Britt's [2011] MD-TRACE model). According to this framework, purposeful reading always involves learners' creating a mental representation of the task and context, which subsequently influences their processing strategies and learning outcomes. Several studies support the notion that relevance instructions are effective in prompting learners to form a task model, which subsequently affects their choice and application of appropriate strategies (e.g., McCrudden et al., 2010; Lehman & Schraw, 2002).

On the basis of these findings from multiple document research, it is reasonable to assume that relevance instructions and guiding questions alter learners' processing mode—either indirectly or directly—from a rather passive reproduction of the content(s) to a more integrative knowledge building and structuring process involving the integration of information from different sources and the generation of new ideas (see also Gil et al., 2010; Wiley & Voss, 1999). Against this background, research on pre-service teachers' KI addressed relevance instructions and guiding questions to examine whether these "low-threshold" instructional means are also effective in fostering the integration of multiple domain-specific knowledge bodies such as CK, PK, and PCK.

Wäschle et al. (2015) and Lehmann et al. (2019) experimentally tested the effect of guiding questions in different reading/writing-based learning settings with three study texts, each pertaining to the CK, PK, or PCK domain. The guiding questions were designed as "strategy activators" (Reigeluth & Stein, 1983, p. 361) to stimulate the integration of information and ideas presented in the text sources. On the basis of a category-driven analysis of participants' written texts, the studies provided converging evidence for the efficacy of guiding questions as regards first-order KI. That is, students in the experimental conditions with guiding questions intentionally attempted to construct relations between domain-specific contents to form new ideas as regards teaching, while students in the control conditions rather tended to process information from different domain-specific sources separately without making novel connections across domains. In another study, Lehmann et al. (2020) performed a computer-linguistic model-based analysis of students' written texts using semantic and structural measures. The results indicated a positive effect of prompting pre-service teachers' integrative structuring of CK, PK, and PCK through guiding questions, thus replicating the prior findings. Interestingly, the studies of Lehmann et al. (2019, 2020) and Wäschle et al. (2015) implemented different writing tasks (essay writing vs. learning journal writing) and study texts on different knowledge domains (mathematical CK/PCK vs. history CK/PCK).

Together they thus indicate that guiding questions can be effectively combined with different types of writing tasks and that their positive effect potentially occurs irrespective of the CK and PCK domains and the topics studied. For the latter conclusion, however, more empirical evidence needs to be generated. What remained largely open in these studies is whether and to what extent a more integrative processing of CK-, PK-, and PCK-specific learning contents affects the integrated use of professional knowledge in application scenarios.

While much of the research on relevance instructions is conducted with a focus on text processing and multiple document comprehension (e.g., McCrudden & Schraw, 2011a, 2011b; McCrudden et al., 2010; Lehman & Schraw, 2002), Zeeb et al. (2019, 2020) conducted a series of experiments to investigate whether a relevance instruction (i.e., an explanation which thoroughly explains why knowledge integration is important) effectively prompts pre-service teachers' KI in a computer-based self-regulated learning environment with two video lectures. Specifically, Zeeb and colleagues examined mathematics and music pre-service teachers' integration of PK and PCK. In one of the experiments, Zeeb et al. (2020) considered whether a more integrative examination of the domain-specific contents presented in the video lectures (as assessed by coding participants' notes and think-aloud protocols) relates to the integrated application of PK and PCK. The results indicated that a (repeated) relevance instruction increased pre-service teachers' use of integrative learning strategies, both cognitive and metacognitive. Also, it was found that the relevance instruction promoted the integrated application of PK and PCK in scenario-based tasks. As regards the relation between the use of integrative strategies during the learning phase and the use of PK and PCK in the application tasks, Zeeb and colleagues found the effect of the relevance instruction on pre-service teachers' integrated knowledge application to be mediated by their use of integrative cognitive strategies during learning. This finding provides evidence for the theoretical conceptualization of (pre-service) teachers' first- and second-order KI (Lehmann, 2020b). However, empirical insights into the influence of relevance instructions on pre-

service teachers' first- and second-order KI in text-based scenarios that involve reading and writing to learn are missing.

The present study

To become a competent teacher, pre-service teachers need to not only acquire but integrate different bodies of their professional knowledge (e.g., Berliner, 2001; Bromme, 2014; Livingston & Borko, 1990). However, the teaching of the different knowledge bodies or domains is often not conducive to pre-service teachers' KI since it is spread across different courses or even institutions (e.g., Darling-Hammond, 2006; Hudson & Zgaga, 2017). Unfortunately, this increases the chance of the knowledge being fragmented into separate memory parts, thus leading to inert knowledge (Renkl et al., 1996). Subsequently, pre-service teachers struggle with adopting multiple domain-specific perspectives on solving tasks or problems (Weinert et al., 1990). To address this issue, recent studies examined instructional prompts such as guiding questions and relevance instructions with promising results (Lehmann et al., 2019, 2020; Wäschle et al., 2015; Zeeb et al., 2019, 2020). However, taking these studies together, the following research questions remain open:

(a) Does a relevance instruction improve pre-service teachers' first-order KI in a text-based learning setting? On the basis of the positive effect of relevance instructions found in the literature on text processing and multiple document comprehension (e.g., McCrudden & Schraw, 2011a; Moreno et al., 2020) and Zeeb et al.'s (2019, 2020) findings from computer-based learning environments, it can be expected that a relevance instruction promotes pre-service teachers' first-order integration of CK, PK, and PCK in a reading- and writing-based learning setting (Hypothesis 1a).

(b) Can the positive effect of guiding questions on first-order KI be replicated with different CK and PCK domains in a corresponding sample? I assume that the positive effect of guiding questions on first-order KI found in prior studies (e.g., Lehmann et al., 2019, 2020; Wäschle et al., 2015) is independent of the subject-matter domain (i.e., CK) and the

associated subject-matter specific didactics (i.e., PCK) to be studied and integrated by pre-service teachers. That is, the effect found previously can be replicated in a learning environment that incorporates different knowledge domains (i.e., German language CK and PCK) with a corresponding sample (German language pre-service teachers). The hypothesis to be tested is that guiding questions promote pre-service German-language teachers' first-order KI of CK, PCK, and general PK when they are learning with multiple domain-specific texts (Hypothesis 1b).

(c) Regarding the two types of prompts, there is also the question of whether their assumed effects are comparable or whether one is superior to the other in enhancing pre-service teachers' first-order KI. I expect guiding questions to be more effective in enhancing first-order KI than a relevance instruction (Hypothesis 2) due to the higher degree of specificity for directing cognitive learning processes accordingly (Davis, 2003; see also Roelle et al., 2015). This argument is further strengthened by the fact that less ambiguity appears to be particularly helpful for learners who are rather inexperienced with the activities demanded by a learning assignment (Linderholm et al., 2014; Rosenshine & Meister, 1992). Besides, the hypothesis is justified by McCrudden et al.'s (2007, 2010) argument that specific prompts better support learners in developing criteria for the evaluation of certain ideas presented in the learning material and are thus more helpful than more general relevance instructions.

(d) To what degree is the integrated application of CK, PK, and PCK (i.e., second-order KI) mediated by an integrative processing of the domain-specific contents during learning (i.e., first-order KI)? The model of teachers' professional competence as a continuum (Blömeke et al., 2015) suggests that any (pre-service) teacher's accomplishment of particular tasks is based on their available knowledge as a cognitive dispositional resource. The incorporation of the conceptualization of knowledge integration as a two-layered construct (first- and second-order KI; Lehmann, 2020b) then specifies that task completion is also

dependent on the degree to which various domain-specific knowledge structures are integrated into a common model (see also Graichen et al., 2019). This corresponds with the theories on learning transfer (e.g., Gick & Holyoak, 1987; Hajian, 2019; Perkins & Salomon, 1992) and leads to the following hypothesis: A more integrated use of CK, PK, and PCK in profession-related application tasks is due to a more integrative learning performance. That is, first-order KI acts as a mediator for second-order KI (Hypothesis 3). The conceptual model underlying the mediation hypothesis is displayed in Figure 1.

** Insert Figure 1 about here **

Method

Participants and design

Eighty-three pre-service elementary school teachers from a German university participated in the study. The participants (86% female; 14% male) had an average age of $M = 24.83$ years ($SD = 2.84$) and a study experience of $M = 7.66$ semesters ($SD = 1.02$). All of them were German language majors and had initial practical teaching experience of $M = 6.40$ months ($SD = 3.96$) through internships in which they had planned and taught an average of $M = 18.94$ lessons ($SD = 16.45$) independently but under supervision. They were native speakers of German (92% first mother tongue; 8% second mother tongue). The participants were recruited in a lecture on learning analysis and evaluation. The pre-service teachers in the lecture could choose to participate in the study or an equivalent alternative activity for course credit. The study incorporated a between-subjects factorial experiment with three parallel groups: (1) control group without prompts (CG; $n = 27$); (2) experimental group with relevance instruction (RI; $n = 28$); (3) experimental group with guiding questions (GQ; $n = 28$). Participants were randomly assigned to one of these groups.

Materials

Learning task and text sources

To stimulate participants' acquisition of teachers' professional knowledge, the present study incorporated a reading- and writing-based learning setting. The setting involved a task sheet which asked them to read three texts, each pertaining to one area of their teacher education studies, and to understand the texts as a whole. Furthermore, participants were instructed to write an essay expressing their thoughts on the topics and their overall understanding. At the end of the task sheet, participants were told that they should take about 60 (max. 75) minutes to read the texts and write their essay. The texts were excerpts from scientific publications and comparable in length and readability (see Table 1). Each domain-specific text enabled the participants to establish connections between information presented in one or both of the other texts. Participants were informed that they could read the texts in any order and also switch back and forth between them while reading and completing the writing task. The study was conducted in the university's test center, which allowed participants to write their essays on a computer. The task sheet and the text sources were provided in hard copy.

** Insert Table 1 about here **

Prompts

Following the example of Zeeb et al. (2019, 2020), the relevance instruction prompt for the RI group started with a description of the different bodies of teacher knowledge (i.e., CK, PK, PCK). Then, it explained by reference to Bromme (2014) why it is important not only to develop comprehensive knowledge bases in these domains but to integrate domain-specific knowledge structures into a common understanding. Furthermore, it clarified that this "merging" of a teacher's knowledge improves their ability to take multiple perspectives in decision making and problem solving and is therefore expected to promote the learning of

their pupils. It concluded with a link to the subsequent writing task by stating that the upcoming learning session would involve three texts, one pertaining to each part of the teacher education program, that should be treated in relation to each other.

Inspired by Wäschle et al. (2015) and Lehmann et al. (2019, 2020), the participants in the GQ group were provided five guiding questions designed to prompt their integration of domain-specific knowledge displayed in multiple documents on CK, PK, and PCK (e.g., "Can you find statements in a text that can be used to explain the content that is presented in the other texts?"; "Can you identify information in the texts that can be linked to conclusions for the design of lessons which are reasonable from multiple perspectives?").

The prompts of both experimental conditions were displayed on the task sheet above (RI) or below (GQ) the writing task. Participants in the control condition received the essay writing task together with a general explanation that each of the texts pertained to one area of their teacher education studies but no prompts. The task sheet, which presented the writing task (and the prompts, where applicable), remained with the participants throughout the phase of reading and essay writing.

Measures

Domain-specific knowledge measures

To assess whether the participants acquired domain-specific knowledge by studying the CK-, PK-, and PCK-text and writing (a more or less integrative) essay, all participants completed three domain-specific knowledge tests. The tests were applied as pre–post measures. Each test consisted of four open short answer and seven closed questions on the contents of the texts. All items were scored dichotomously as correct (1 point) or incorrect (0 points). Hence, the maximum score for each test was eleven. The items were developed on the basis of the text sources. Each test was reviewed by an expert of the particular domain for the sake of face validity. The tests assessed both factual knowledge and conceptual understanding (see Table 2 for sample items).

I used the Kuder-Richardson Formula 20 (KR20; Kuder & Richardson, 1937), which is an equivalent to Cronbach's alpha for dichotomous items, to determine the internal consistencies of the knowledge tests. The KR20 coefficients of all three tests were lower than the 0.7 value (see Table 2), which is undesirable according to the literature (e.g., Thompson, 2010). This may result from the fact that the knowledge tests were composed to capture the breadth of prior and acquired knowledge. The tests therefore queried different facets of the topics covered in the texts and were not optimized for high homogeneity. Moreover, the rather low reliabilities may result from the fact that items assessed both factual knowledge and conceptual understanding. Since the intention of the pre–post knowledge assessment was to compare the gains in CK, PK, and PCK across the experimental conditions and because of the high content validity (all items were deduced directly from the text sources), the lower KR20 values should not be a severe issue.

** Insert Table 2 about here **

First-order KI (integrative learning performance measures)

To assess participants' first-order KI, I conducted a category-driven content analysis on their essays. The analysis followed a procedure which has been previously used in research on multiple-document comprehension (e.g., Gil et al., 2010; Wiley & Voss, 1999) and adapted to pre-service teachers' learning across different domains (e.g., Lehmann et al., 2019, Lehmann, 2020a). The procedure involved identifying the nature of pre-service teachers' learning as either separative or integrative. Separative learning refers to the processing, acquisition, and organization of information and knowledge while making no connections to other disciplines, domains of knowledge, and/or topics. By contrast, integrative learning refers to constructive modes of processing that target first-order KI by interrelating and

merging originally unconnected domain-specific knowledge entities into a common mental model (e.g., Lehmann, 2020b; Lee & Turner, 2017).

To score participants' learning performance, all essays were first parsed into idea units. Then, each idea unit was coded to indicate whether it was (a) a rather passive reproduction of what was written in the text source (*borrowing*), (b) transformed within one of the knowledge domains by combining information from a text with further information from the same text or from prior knowledge (*elaboration within domain*), or (c) elaborated across domains by relating information from one text with information from either one or both other texts, potentially with added information from prior knowledge (*elaboration across domains*). While (a) and (b) are indicators of separative learning, (c) refers to integrative learning. Following Britt and Sommer (2004) and Gil et al. (2010), I treated the frequency of *switches* between the three knowledge domains CK, PK, and PCK as another indicator of integrative learning. If an essay contained 15 idea units in total and the first six units pertained to the CK domain, the next six pertained to the PK domain, and the last seven pertained to the PCK source, this would be coded as two switches and indicate only marginal integration. By contrast, an essay that switches from one to another domain every one or two idea units displays a high degree of interconnectedness, thus indicating considerable efforts on integrative learning. To estimate participants' first-order KI, I calculated an overall index by summing the scores in the integrative learning measures "elaboration across domains" and "switches."

Idea units that presented information not addressed in the text sources were coded as *additions.* Additions occurred when a student included further information from prior knowledge, stated their personal opinion, or expressed metacognitive thoughts. The coding scheme is displayed in Table 3. It was applied by two raters who were familiar with the text sources but blind to the experimental conditions. The raters segmented and coded the same random subset of data (30%) independently with satisfying interrater agreement (Cohen's

κ = .75). Disagreements were settled through discussion between the two raters. After consensus was reached, each rater coded half of the remaining data.

** Insert Table 3 about here **

Second-order KI (integrated knowledge application test)

Assessing participants' second-order KI involved the justified evaluation and improvement of a tabular lesson plan. Specifically, the integrated knowledge application test asked the participants to evaluate how the planned lesson was (not) appropriate for promoting spelling skills in a third-grade class. The intended learning outcomes were noted above the tabular outline of the lesson (e.g., "The pupils … understand that there are a and au words related to ä and äu words; … realize that the related words are a solution aid for distinguishing between ä and e or between äu and eu words; … put the solution aid (umlaut rule; morphological principle) into their own words and practice"). The lesson plan incorporated time, classroom activities, method/class arrangement, and material as planning dimensions. The lesson was designed to include both strengths and weaknesses that could be identified on the basis of the three domain-specific text sources. The evaluation by the participants was pre-structured in that they were instructed to "(a) list positive features of the draft ('What is good about the planned lesson?') and justify why and, if necessary, under what conditions a feature indicated successful learning and (b) make suggestions for improvement ('What is not well-designed, could therefore be improved, and how?')." A corresponding evaluation template allowed participants to fill in the positive features they identified and those that needed improvement in a pre-structured table together with their reasoning behind each.

Two independent raters coded participants' responses (i.e., each feature with its associated rationale). Responses were first segmented into idea units and then coded for

whether or not they referred to the previously read domain-specific texts and whether information from the texts had been correctly applied or misinterpreted. In accordance with Graichen et al. (2019), each reference to one of the domain-specific texts was coded with the respective category (*CK application, PK application, PCK application, incorrect knowledge application*). Reasonable ideas that did not refer to any of the texts were coded as *pre-knowledge application.* Ill-founded ideas without text reference or misconceptions were coded as *incorrect knowledge application*. Two independent raters coded the same random subset (30%) of the data. With intraclass correlation coefficients (ICCs) between .82 and .94, the interrater reliability can be considered excellent (Cicchetti, 1994; see Table 4 for examples of responses and all ICCs). Disagreements between raters were resolved through discussion. Then, each rater coded half of the remaining data.

For an estimation of participants' *second-order KI*, the following scoring scheme was applied on the feature level after the raters coded the data. If the rationale associated with a feature solely included one or more ideas that were coded as incorrect knowledge application, no points were given. If a named feature was justified with only CK, PK, PCK, or prior knowledge, one point was awarded. This also applied to rationales that included multiple references to a single domain. If the rationale incorporated the application of knowledge from two distinct domains (e.g., CK and PK), two points were given. Integration of knowledge from all three domains (i.e., CK, PK, and PCK) into a feature's rationale scored three points. If prior knowledge was integrated additionally to the use of CK, PK, and/or PCK, an additional point was given (e.g., the use of CK, PK, and prior knowledge was worth three points). One point was deducted from a rationale's score for each idea unit that represented an incorrect knowledge application. That is, a rationale that consisted of CK, PK, and PCK (three points) and an incorrect knowledge application (minus one point) was worth only two points. Finally, I calculated an overall index for participants' second-order KI by adding up the points scored for each feature with its associated rationale.

** Insert Table 4 about here **

Procedure

The study took place in two sessions. The first session was conducted online. Participants received a link to an online survey, which comprised a demographic data survey including a code-generating item and the domain-specific knowledge tests (pre-test). Then, the participants received a reminder about the second session of study, which took place in a large test center during the regular course time approximately ten days after the first session. In the laboratory session, participants were first randomly assigned to the three experimental conditions. Then, they completed a short form of the demographic data survey with the same code-generating item so that the data of the two sessions could be matched. Next, they received the task sheet with the writing task and the instructional scaffolds (i.e., relevance instruction, guiding questions), where applicable, as well as the text sources described in Table 1. On average, it took the participants 58 minutes ($SD = 7.66$) to read the texts and complete the writing task. After the participants had submitted their essays on the computer and the text sources to the experimenter, they received a post-test booklet including the same domain-specific knowledge tests and the integrated knowledge application test.

Data analysis

I conducted power analyses using the G*Power software (version 3.1.9.7) to determine whether the sample size was sufficient to detect the effect of each test at an alpha-level of .05. Assuming a large effect ($\eta^2 = .14$), which seems reasonable in view of the studies on the effectiveness of relevance instructions (Zeeb et al., 2020) and guiding questions (e.g., Lehmann et al., 2019), and a desired power of .80, a sample size of 21 subjects per experimental group (63 subjects in total) was required for the ANOVAs to yield significant results. As to the mediation hypothesis, I conducted a power analysis using the joint

significance test described by MacKinnon et al. (2002). Assuming a medium mediation effect (Graichen et al., 2019), the joint significance test with a power of .80 revealed that a sample of $n = 74$ was enough to yield a significant result. This is in line with the recommendation given by Fritz and MacKinnon (2007).

I tested within-subject effects and group differences for significance using (multivariate) analyses of variance ([M]ANOVA) with the experimental condition (CG, RI, GQ) as a between-subjects factor and with repeated measures, where applicable. Tukey's honestly significant difference (HSD) test was used for post-hoc pairwise comparisons. Partial eta squared (η_p^2) was used as an effect size measure. In cases where the homogeneity of variance assumption was not met, Welch's test was applied. With reference to Cohen (1988), η_p^2-coefficients $< .06$ were interpreted as a small effect, between .06 and .13 as a medium effect, and $> .13$ as a large effect. Mediation analysis was used as a means of examining the relationships between experimental condition and first- and second-order KI. All analyses were performed with SPSS 26. Mediation models were estimated with the PROCESS 4.0 macro for SPSS provided by Hayes (2022) with $m = 5,000$ bootstrap samples. Effects were only considered significant if the confidence interval did not include zero.

Results

Initial data analysis

As an initial data analysis, I tested for differences between experimental groups regarding study experience (semester), practical teaching experience (in terms of the duration of internships and having planned and held lessons independently but under supervision), time on task, the total amount of idea units included in participants' essays, and the quantity of features identified as part of the integrated knowledge application test. No significant differences were found (all Fs ≤ 2.074, ps $\geq .132$). Moreover, the experimental groups did not differ in their domain-specific pre-knowledge (all Fs ≤ 1.112, ps $\geq .334$). Thus, these

variables did not influence the results. Table 5 provides an overview of the descriptive and inferential statistical results for these variables.

** Insert Table 5 about here **

Domain-specific knowledge acquisition

A repeated-measures MANOVA with the scores in the domain-specific pre- and post-knowledge tests as a dependent measure and the experimental condition as an independent measure assessed the domain-specific knowledge acquisition that resulted from the learning session in the experiment. The results indicated a significant increase in domain-specific knowledge (see Table 6 for descriptive statistics), Wilks' λ = .031, $F(3,78)$ = 815.2, $p < .001$, η_p^2 = .969 (strong effect), but no effect of experimental condition, Wilks' λ = .940, $F(6,156)$ = .818, p = .557, η_p^2 = .031, and no significant interaction, Wilks' λ = .980, $F(6,156)$ = .270, p = .950, η_p^2 = .010. To identify the domains in which participants' knowledge gains were significant, I conducted a repeated-measures ANOVA for each domain. The results showed that participants' knowledge gains were significant with strong effect sizes in all three domains (CK: Wilks' λ = .077, $F(1,80)$ = 954.0, $p < .001$, η_p^2 = .923; PK: Wilks' λ =.121, $F(1,80)$ = 583.1, $p < .001$, η_p^2 = .879; PCK: Wilks' λ = .115, $F(1,80)$ = 616.7, $p < .001$, η_p^2 = .885). Since there were no significant interactions between knowledge test scores and experimental condition, the observed increase in domain-specific knowledge was comparable across conditions (CK: Wilks' λ = .991, $F(2,80)$ = .359, p = .700, η_p^2 = .009; PK: Wilks' λ = .996, $F(2,80)$ = .179, p = .837, η_p^2 = .004; PCK: Wilks' λ = .994, $F(2,80)$ = .230, p = .795, η_p^2 = .006).

** Insert Table 6 about here **

Effects on first-order KI

First-order KI was estimated by using the following coding categories as dependent measures of a separative or integrative processing of the domain-specific learning content. Separative learning included (a) borrowings and (b) elaborations within domains. Integrative learning was assessed by (c) elaborations across domains and (d) the number of switches between domain-specific statements to account for the merging of information in participants' essays, together forming an index for first-order KI. Descriptive statistics for these dependent variables and for additions, which were not captured by the separative and integrative learning measures, are reported in Table 7.

** Insert Table 7 about here **

A MANOVA with the different measures indicated that the overall learning performance was affected by the experimental condition, Wilks' $\lambda = .535$, $F(10,152) = 5.584$, $p < .001$, $\eta_p^2 = .269$ (strong effect). Follow-up ANOVAs, with *Welch's F* for the variables with heterogeneous variances, estimated the effects on each performance measure separately. For the separative learning measures, the results showed a significant effect of experimental condition on borrowings in essays, *Welch's F*$(2,52.26) = 9.654$, $p < .001$, $\eta_p^2 = .213$ (strong effect). Tukey's HSD indicated that participants in the RI and GQ group included significantly fewer borrowings from the text sources ($M_{RI} = 8.75$; $M_{GQ} = 7.68$) than participants in the control group ($M_{CG} = 12.93$). The difference in borrowings between RI and GQ was not significant. No significant differences were found for elaborations within domains, $F(2,80) = 2.028$, $p = .138$, $\eta_p^2 = .048$, or for additions, $F(2,80) = .582$, $p = .561$, $\eta_p^2 = .014$.

For the integrative learning performance measures, the results revealed strong effects of the experimental condition on elaborations across domains, *Welch's F*$(2,51.63) = 25.23$,

$p < .001$, $\eta_p^2 = .354$, and switches, *Welch's F*(2,51.23) = 9.038, $p = .001$, $\eta_p^2 = .152$. The post-hoc analysis showed that participants in the RI and GQ condition included more elaborations across domains ($M_{RI} = 5.68$; $M_{GQ} = 7.46$) and made more switches ($M_{RI} = 10.36$; $M_{GQ} = 11.46$) than the control group ($M_{CG\_elaborations-across-domains} = 2.85$ and $M_{CG\_switches} = 6.89$). Regarding the differences between the RI and GQ conditions, the post-hoc analysis indicated that the GQ group generated significantly more elaborations across domains than the RI group, but the difference in switches failed to meet statistical significance.

Together, these results provide evidence for Hypotheses 1a and 1b in that both types of prompts, the relevance instruction and guiding questions, promote pre-service teachers' first-order KI in learning with multiple domain-specific texts. Interestingly, the guiding questions provoked more integrative elaborations across domains than the relevance instruction (which partially supports Hypothesis 2). However, there was no significant difference for switching between different knowledge domains in essay writing.

Effects on second-order KI (mediation analysis)

Second-order KI was estimated under consideration of participants' CK, PK, and PCK application in evaluating and improving a lesson plan. The descriptive statistics for all knowledge application measures and for second-order KI are displayed in Table 8.

** Insert Table 8 about here **

To investigate whether an integrative processing of the domain-specific contents affects the integrated use of knowledge (Hypothesis 3), I conducted a mediation analysis with the experimental condition as the independent variable (X), the first-order KI index as the mediator (M), and the second-order KI index as the dependent variable (Y) (see Figure 1). This allowed me to examine whether the participants' integrated application of domain-

specific knowledge (i.e., second-order KI) was dependent on their integrative learning performance (i.e., first-order KI) in the reading- and writing-based learning setting.

Results showed a significant direct effect of experimental condition on participants' second-order KI, $F(2,80) = 7.951$, $p < .001$ ($c_{RI} = 3.909$, $p = .002$, 95%CI_{RI} [1.51, 6.31], $c_{GQ} = 4.409$, $p < .001$, 95%CI_{GQ} [2.01, 6.81]). Hence, both experimental conditions improved participants' integrated application of CK, PK, and PCK compared to the control condition. In addition, the results suggested that the experimental conditions involving an instructional scaffold enhanced first-order KI in participants compared to the control condition ($a_{RI} = 6.295$ and $a_{GQ} = 9.188$), thus providing further evidence for Hypotheses 1a and 1b (see Table 9). More importantly, the mediation analysis showed that participants who were more successful at first-order KI achieved higher scores on the integrated knowledge application test ($b = .353$), with no significant direct effect of the experimental condition on second-order KI remaining ($ps > .129$). A bootstrap confidence interval for the indirect effect based on 5,000 bootstrap samples did not include zero (95%CI_{RI} [1.07, 3.56]; 95%CI_{GQ} [1.69, 5.03]). Hence, the results exhibited evidence for Hypothesis 3. That is, the more pre-service teachers engaged in first-order KI when learning with multiple domain-specific texts, the more they integrated knowledge from multiple domains (i.e., CK, PK, PCK) in evaluating and improving a lesson plan. Table 8 and Figure 2 summarize the results of the mediation analysis. Contrasting the effects of the scaffolded conditions RI and GQ suggested no significant differences between the two experimental groups with prompts ($ps > .349$).

** Insert Table 9 and Figure 2 about here **

Discussion

In the present study, I examined the effects of a relevance instruction and of guiding questions on pre-service teachers' first- and second-order KI in a reading- and writing-based

learning setting with multiple domain-specific texts. I aimed at identifying instructional prompts that are effective in promoting pre-service teachers' KI as a constructive form of integrative knowledge building and structuring (i.e., first-order KI). Another objective was to test the assumption that first-order KI mediates second-order KI, the latter referring to the simultaneous (integrated) use of knowledge from diverse domains (i.e., CK, PK, PCK) in profession-related application tasks such as lesson planning.

The study provides evidence for the efficacy of a relevance instruction and guiding questions for pre-service teachers' first-order KI from multiple texts. Specifically, it was found that both types of prompts stimulated pre-service teachers to generate and provide integrative elaborations (which involve mental interrelations across multiple domains) and to merge domain-specific ideas by switching back and forth between domains more frequently in their essays. This higher engagement in cognitive processes related to the concept of first-order KI was accompanied by less borrowings in the essays. Pre-service teachers who received the prompts thus changed their strategic processing of the domain-specific learning contents from a more summarizing approach to an integrative knowledge building and structuring across domains.

The present results can also be interpreted in terms of the idea that learners who are confronted with multiple documents first construct a (more or less beneficial) task model on the basis of the instructions given (e.g., by the reading/writing task and prompts) and their deduced reading goal (Britt et al., 2017). The task model then guides learners' decisions and actions in terms of focusing and integratively elaborating domain-specific information and ideas across domains. In this study, both types of prompts, relevance instructions and guiding questions, helped pre-service teachers to better allocate their attention and information processing. This allocation of focus and information processing was aimed at achieving a comprehensive understanding of the domain-specific learning contents as a whole, facilitating the integration of CK, PK, and PCK. The effectiveness was evidenced by increased efforts in

integrative elaborations and switches, as well as less borrowings. It is important to note that domain-specific knowledge acquisition was not affected by this, as the knowledge gains were significant for all domains in all experimental conditions and did not differ between groups. These findings are in line with prior studies (e.g., Lehmann et al., 2019, 2020; Wäschle et al., 2015; Zeeb et al., 2019) and add to the body of research on the effectiveness of relevance instructions and guiding questions in promoting pre-service teachers' first-order KI. Additionally, the present study extends prior findings in that it involved another subject-matter domain (i.e., German language) and combined a relevance instruction with pre-service teachers' reading/writing-based learning with multiple domain-specific texts.

Furthermore, the study highlights how important first-order KI is for second-order KI in lesson planning, which is congruent with the theoretical conceptualization of first- and second-order KI (Lehmann, 2020b) and its association with Blömeke et al.'s (2015) model of teachers' professional competence, as well as theories on learning transfer (e.g., Gick & Holyoak, 1987; Hajian, 2019; Perkins & Salomon, 1992). Pre-service teachers who dealt with the learning contents under consideration of a relevance instruction or guiding questions performed better not only in first-order KI but also in the integrated application of CK, PK, and PCK. Obviously, pre-service teachers' first-order KI mediated their second-order KI in evaluating and improving a worked-out lesson plan. No significant differences were found between the two types of prompts with regard to second-order KI. However, contrasting the effects of a relevance instruction and guiding questions in an analysis of the first-order KI measures drew an interesting picture: The descriptive statistics clearly indicated a tendency toward the more specific guiding questions for both integrative learning measures (i.e., elaborations across domains and switches) and for participants' overall first-order KI. However, only the difference in pre-service teachers' integrative elaborations across domains proved significant (which partly supports Hypothesis 2). The increased merging of domain-specific information due to switching between domains in the essays was not significant.

Hence, there is only partial evidence that guiding questions are more beneficial for fostering first-order KI in pre-service teachers as compared to a relevance instruction. These mixed results demand potential explanations.

First, it is surprising that the guiding questions were not (more) superior in enhancing pre-service teachers' first-order KI, because university students in general and pre-service teachers in particular have been found repeatedly to rarely engage in or struggle with integrative learning processes if not specifically stimulated and assisted in doing so (e.g., Lehmann et al., 2019; Gil et al., 2010; Wäschle et al., 2015). Pre-service teachers can thus be regarded as rather inexperienced in effectively applying strategies that target the integration of CK, PK, and PCK. This makes directed prompts such as guiding questions appear to be more effective than a relevance instruction because they are more specific in eliciting the cognitive processes relevant to first-order KI, which is in line with prior studies on the role of specificity in instructional scaffolds (e.g., Roelle et al., 2015). On the other hand, specificity can be detrimental to learning if the scaffold directs thought processes that interfere with the personal strategies that have already been developed. Students are then confused by the prompt to perform a specific (mental) action within a particular learning activity (e.g., in a reading- and writing-based setting) if it does not fit their personal approach. This interpretation is not only rational from a theoretical perspective; it also finds empirical support in that the variance as regards both integrative learning measures and the first-order KI index score was larger in the guiding questions condition. Future work could therefore address the personalized and adaptive realization of prompts for knowledge integration, albeit requiring automated assessment and feedback, which is an important field of research in itself.

Another explanation may lie in the two integrative learning performance (sub-)measures. Although switches were considered a valid integration measure in previous research (e.g., Britt & Sommer, 2004; Gil et al., 2010), they simply indicate the merging of

domain-specific knowledge entities or information. Elaborations across domains, on the other hand, demand the identification of domain-specific knowledge entities that can be combined across domains to form a coherent idea (e.g., Lehmann et al., 2019; Graichen et al., 2019; Wäschle et al., 2015). Beyond that, pre-service teachers needed to verbally express the elaborations across domains in their essays. Overall, this is certainly more challenging. Hence, one could argue with regard to the integrative learning measures that including elaborations across domains in a written account is more valuable for knowledge integration than switching back and forth between different knowledge domains. For the comparison of the two prompts, this means that the effect of guiding questions on elaborations across domains found in the study is more important than the missing effect on switches.

Pedagogical implications

With respect to pedagogical implications for pre-service teacher education, it is possible to derive several suggestions from the present study. First, the findings encourage the use of guiding questions and/or relevance instructions for supporting pre-service teachers' integrative learning of CK, PK, and PCK as opposed to assuming that they will develop a well-integrated knowledge base across domains in a purely self-regulated manner when confronted with various domain-specific learning material. Second, more specific integration prompts such as guiding questions facilitate pre-service teachers' integrative elaboration of domain-specific ideas better than more general relevance instructions. Finally, the implementation of such scaffolds affects pre-service teachers' ability to integrate multiple domain-specific perspectives into application tasks such as lesson planning. However, since this effect is mediated by pre-service teachers' first-order KI, it appears important for instructors to embed integrative learning activities such as identifying, evaluating, discussing, and elaborating information and ideas across the conceptual border of certain knowledge domains (CK, PK, PCK) into course routines to train first-order KI as a particular learning strategy.

Limitations of the study and future directions

As with all research, there are several constraints to this study that need to be addressed. First, the operationalization of first-order KI as a summative index score needs to be critically discussed. While the underlying integrative learning measures themselves have sufficient inter-rater reliability and were considered valid in several prior studies (e.g., Britt & Sommer, 2004; Gil et al., 2010; Wiley & Voss, 1999), it remains open to what degree combining these measures into an estimate of a larger concept like first-order KI maintains sufficient construct validity. Likewise, the construct validity of second-order KI might be an issue. For this, the present study modified Graichen et al.'s (2019), approach of scoring and summing references to domain-specific learning materials (i.e., text sources) in a knowledge application test for determining the integrated use of CK, PK, and PCK. However, it is well known that validation is not an activity that occurs once assessments are developed; rather, it is an ongoing process. In light of these concerns, I therefore suggest that future studies draw on complementary approaches for the assessment of pre-service teachers' first- and second-order KI.

An estimate of first-order KI might involve more diverse integrative learning measures alongside the elaborations across domains and switches. For example, instructing students to try to benefit from external learning strategies such as text-highlighting and annotating (since this has been found to foster readers' integrated understanding of multiple texts; e.g., Leroy et al., 2020; Kobayashi, 2009) could be fruitful in two ways: (1) An analysis of pre-service teachers' text-highlights and annotations of the domain-specific text sources could supplement the integrative learning measures of the present study and thus contribute to a methodologically sound estimate of first-order KI. Thereby, the text-highlights could also be used in retrospective interviews with the participants on why they perceived particular statements or information (i.e., pieces of knowledge) to be relevant for the integration of CK, PK, and PCK. The reasons given would provide even more insight into pre-service teachers'

first-order KI. (2) The external strategies are worth investigating in regard to whether they improve pre-service teachers' integration of CK, PK, and PCK when they are learning from multiple domain-specific texts.

Another limitation of the present study is that the time constraints imposed on the learning process and the controlled laboratory setting may restrict the scope of the findings and pose a potential threat to ecological validity. While the study is in line with previous research on the effectiveness of various prompts in enhancing pre-service teachers' first- and second-order KI (Lehmann et al., 2019, 2020; Wäschle et al., 2015; Zeeb et al., 2019, 2020), it focused solely on the immediate effects within a reading- and writing-based learning environment. Consequently, it did not examine potential mid- and long-term effects on pre-service teachers' KI. Therefore, future studies should investigate the nature of first- and second-order KI and the impact of different prompts using delayed testing and real-world settings. For instance, prompts could be incorporated into reading and writing assignments within a regular course on PCK (which can be depicted as a connector between CK and PK).

Furthermore, the participants were regarded as being rather inexperienced in using integrative learning strategies. Accordingly, the guiding questions were designed to direct specific mental activities related to first-order KI. However, such directed prompts also have drawbacks, especially when they are not tailored to the level of individuals' expertise and their personal learning strategies (Rosenshine & Meister, 1992). Hence, future studies may consider the degree to which participants generally process the learning contents from the different knowledge domains (CK, PK, PCK) in an integrative and separative manner as potentially confounding variables.

What limits the generalizability of the present findings is that they are based on a sample of pre-service primary teachers of German. Future research should expand the scope to include pre-service teachers who will teach different subjects later on in their career and intend to work, for example, in secondary schools. Additionally, it is crucial to note that the

text documents utilized as learning resources in this study should be considered exemplary. While future research should encompass documents covering alternative topics, it is important to highlight that the current study has already replicated prior findings on the efficacy of prompts (e.g., Lehmann et al., 2019; Wäschle et al., 2015) with different CK and PCK domains, as well as different topics within each domain.

Last, the present study examined the integrated application of CK, PK, and PCK with regard to the evaluation and improvement of a completed lesson draft. Hence, the finding on the importance of first-order KI for second-order KI was limited to lesson planning. Further research should therefore examine to what degree an improved first-order KI affects other profession-related tasks (e.g., the evaluation and design of learning tasks and material) as well as the actual implementation of a lesson plan and classroom practice.

Conclusion

In spite of its limitations, the study certainly adds to our understanding of pre-service teachers' knowledge integration and how to support it. As regards theory, the findings strengthen the conceptualization of (pre-service) teachers' knowledge integration of multiple domains as a two-layered construct: (1) as a form of cognitive-constructive learning that interrelates, connects, and merges originally unconnected entities of CK, PK, and PCK into a more coherent knowledge structure (first-order KI); and (2) as a form of integrated, simultaneous application of domain-specific knowledge (second-order KI). Concerning pre-service teacher education, the findings of the study show that a relevance instruction and guiding questions are effective means of promoting pre-service teachers' integration of CK, PK, and PCK in both learning and application.

References

Bannert, M. (2009). Promoting self-regulated learning through prompts. *Zeitschrift für Pädagogische Psychologie, 23*(2), 139-145. https://doi.org/10.1024/1010-0652.23.2.139

Barzel, B., Eichler, A., Holzäpfel, L., Leuders, T., Maaß, K., & Wittmann, G. (2016). Vernetzte Kompetenzen statt trägen Wissens – Ein Studienmodell zur konsequenten Vernetzung von Fachwissenschaft, Fachdidaktik und Schulpraxis. In A. Hoppenbrock, R. Biehler, R. Hochmuth & H.-G. Rück (Eds.), *Lehren und Lernen von Mathematik in der Studieneingangsphase* (pp. 33–50). Wiesbaden: Springer. https://doi.org/10.1007/978-3-658-10261-6_3

Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft, 9*(4), 469–520. https://doi.org/10.1007/s11618-006-0165-2

Berliner, D. C. (2001). Learning about and learning from expert teachers. *Educational Research, 35*(5), 463–482. https://doi.org/10.1016/S0883-0355(02)00004-6

Bigot, L. L., & Rouet, J.-F. (2007). The impact of presentation format, task assignment, and prior knowledge on students' comprehension of multiple online documents. *Journal of Literacy Research, 39*(4), 445–470. https://doi.org/10.1080/10862960701675317

Björnsson, C. H. (1983). Readability of newspapers in 11 languages. *Reading Research Quarterly, 18*(4), 480–497. https://doi.org/10.2307/747382

Blömeke, S., Gustafsson, J. E., & Shavelson, R. J. (2015). Beyond dichotomies. Competence viewed as a continuum. *Zeitschrift für Psychologie, 223*(1), 3–13. https://doi.org/10.1027/2151-2604/a000194

Bromme, R. (2014). *Der Lehrer als Experte: Zur Psychologie des professionellen Wissens.* Münster: Waxmann.

Britt, M. A., Rouet, J.-F., Durik, A. M. (2017). *Literacy beyond text comprehension. A theory of purposeful reading.* New York: Routledge. https://doi.org/10.4324/9781315682860

Britt, M. A., & Sommer, J. (2004). Facilitating textual integration with macro structure-focusing tasks. *Reading Psychology, 25*(4), 313–339. https://doi.org/10.1080/02702710490522658

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*(4), 284–290. https://doi.org/10.1037/1040-3590.6.4.284

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Darling-Hammond, L. (2006). Constructing 21st-Century teacher education. *Journal of Teacher Education*, *57*(3), 300–314. https://doi.org/10.1177/0022487105285962

Davis, E. (2003). Prompting middle school science students for productive reflection: Generic and directed prompts. *Journal of the Learning Sciences, 12*(1), 91–142. https://doi.org/10.1207/S15327809JLS1201_4

Fritz, M.S., & MacKinnon, D.P. (2007). Required sample size to detect the mediated effect. *Psychological Science, 18*, 233–239. https://doi.org/10.1111/j.1467-9280.2007.01882.x

Gick, M. L., & Holyoak, K. J. (1987). The cognitive basis of knowledge transfer. In S. M. Cormier & J. D. Hagman (Eds.), *Transfer of learning: Contemporary research and applications* (pp. 9–46). Academic Press.

Gil, L., Bråten, I., Vidal-Abarca, E., & Strømsø, H. I. (2010). Summary versus argument tasks when working with multiple documents: Which is better for whom? *Contemporary Educational Psychology, 35*(3), 157–173. https://doi.org/10.1016/j.cedpsych.2009.11.002

Gottein, H.-P. (2020). Closing the Gap: An Innovative Learning Environment for Enabling Pre-Service Teachers to Put Theoretical Knowledge into Action. In T. Lehmann (Ed.), *International Perspectives on Knowledge Integration: Theory, Research, and Good Practice in Pre-service Teacher and Higher Education* (pp. 231–254). Leiden: Brill | Sense. https://doi.org/10.1163/9789004429499_011

Graichen, M., Wegner, E., & Nückles, M. (2019). Wie können Lehramtsstudierende beim Lernen durch Schreiben von Lernprotokollen unterstützt werden, dass die Kohärenz und Anwendbarkeit des erworbenen Professionswissens verbessert wird? *Unterrichtswissenschaft, 47,* 7–28. https://doi.org/10.1007/s42010-019-00042-x

Hajian, S. (2019). Transfer of learning and teaching: A review of transfer theories and effective instructional practices. *IAFOR Journal of Education*, *7*(1), 93-111. https://doi.org/10.22492/ije.7.1.06

Hanisch, A.-K. (2018). *Kognitive Aktivierung im Rechtschreibunterricht. Eine quasi-experimentelle Interventionsstudie in der Grundschule*. Münster: Waxmann.

Harr, N., Eichler, A., & Renkl, A. (2014). Integrating pedagogical content knowledge and pedagogical/psychological knowledge in mathematics. *Frontiers in Psychology, 5,* 1–10. https://doi.org/10.3389/fpsyg.2014.00924

Hayes, A. F. (2022). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. 3rd Edition. New York: The Guilford Press.

Hudson, B., & Zgaga, P. (2017). History, context and overview: Implications for teacher education policy, practice and future research. In B. Hudson (Ed.), *Overcoming fragmentation in teacher education policy and practice* (pp. 1–26). Cambridge: Cambridge University Press.

Janssen, N., & Lazonder, A. W. (2016). Supporting pre-service teachers in designing technology-infused lesson plans. *Journal of Computer Assisted Learning, 32*(5), 456–467. https://doi.org/10.1111/jcal.12146

Kobayashi, K. (2009). Comprehension of relations among controversial texts: Effects of external strategy use. *Instructional Science, 37*(4), 311–324. https://doi.org/10.1007/s11251-007-9041-6

König, J. (2010). Lehrerprofessionalität. Konzepte und Ergebnisse der internationalen und deutschen Forschung am Beispiel fachübergreifender, pädagogischer Kompetenzen. In J.

König & D. B. Hofmann (Eds), *Professionalität von Lehrkräften. Was sollen Lehrkräfte im Lese- und Schreibunterricht wissen und können?* (pp. 40-105). Berlin: DGLS.

Krauss, S. (2011). Das Experten-Paradigma in der Forschung zum Lehrerberuf. In E. Terhart, H. Bennewitz & M. Rothland (Eds.), *Handbuch der Forschung zum Lehrerberuf* (pp. 186–191). Münster: Waxmann.

Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika, 2*(3), 151–160. https://doi.org/10.1007/BF02288391

Lachner, A., Jarodzka, H., & Nückles, M. (2016). What makes an expert teacher? Investigating teachers' professional vision and discourse abilities. *Instructional Science, 44*(3), 197–203. https://doi.org/10.1007/s11251-016-9376-y

Lee, J., & Turner, J. E. (2017). Extensive knowledge integration strategies in pre-service teachers: The role of perceived instrumentality, motivation, and self-regulation. *Educational Studies.* https://doi.org/10.1080/03055698.2017.1382327

Lehman, S., & Schraw, G. (2002). Effects of coherence and relevance on shallow and deep text processing. *Journal of Educational Psychology, 94*(4), 738–750. https://doi.org/10.1037/0022-0663.94.4.738

Lehmann, T. (2020a). Investigating the relation between pre-service teachers' argumentative writing and knowledge integration when learning with multiple texts. *Journal of Psychological and Educational Research, 28*(2), 75–106.

Lehmann, T. (2020b). What is Knowledge Integration of Multiple Domains and How Does it Relate to Teachers' Professional Competence? In T. Lehmann (Ed.), *International Perspectives on Knowledge Integration: Theory, Research, and Good Practice in Pre-service Teacher and Higher Education* (pp. 9–29). Leiden: Brill | Sense. https://doi.org/10.1163/9789004429499_002

Lehmann, T. (2022). Student teachers' knowledge integration across conceptual borders: the role of study approaches, learning strategies, beliefs, and motivation. *European Journal*

of Psychology of Education, 37(4), 1189–1216. https://doi.org/10.1007/s10212-021-00577-7

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

Lehmann, T., Pirnay-Dummer, P., & Schmidt-Borcherding, F. (2020). Fostering integrated mental models of different professional knowledge domains: Instructional approaches and model-based analyses. *Educational Technology Research and Development, 68*(3), 905–927. https://doi.org/10.1007/s11423-019-09704-0

Lenhard, W., & Lenhard, A. (2017). *Berechnung des Lesbarkeitsindex LIX nach Björnsson*. Online available: http://www.psychometrica.de/lix.html. Bibergau: Psychometrica. https://doi.org/10.13140/RG.2.1.1512.3447

Leroy, C., Gerjets, P., Oestermeier, U., Kammerer, Y. (2020). Simultaneous presentation of multiple documents and text-highlighting: Online integrative processes and offline integrated understanding. *Scientific Studies of Reading, 2*(25), 179–192. https://doi.org/10.1080/10888438.2020.1784903

Lilliedahl, J., Wikman, T., & Hansén, S. (2020). Coherence within Teacher Education: Swedish Lessons. In T. Lehmann (Ed.), *International Perspectives on Knowledge Integration: Theory, Research, and Good Practice in Pre-service Teacher and Higher Education* (pp. 291–312). Leiden: Brill | Sense. https://doi.org/10.1163/9789004429499_013

Linderholm, T., Therriault, D. J., & Kwon, H. (2014). Multiple science text processing: Building comprehension skills for college student readers. *Reading Psychology*, *35*(4), 332–356. https://doi.org/10.1080/02702711.2012.726696

Lipowsky, F. (2020). Unterricht. In E. Wild & J. Möller (Eds.), *Pädagogische Psychologie* (pp. 69–105). Berlin: Springer. https://doi.org/10.1007/978-3-662-61403-7_4

Livingston, C., & Borko, H. (1990). High school mathematics review lessons: Expert-novice distinctions. *Journal of Research in Mathematics Education, 21*(5), 372–387. https://doi.org/10.2307/749395

MacKinnon, D.P., Lockwood, C.M., Hoffman, J.M., West, S.G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods, 7,* 83–103. https://doi.org/10.1037/1082-989X.7.1.83

Mateos, M., Martín, E., Cuevas, I., Villalón, R., Martínez, I., & González-Lamas, J. (2018). Improving written argumentative synthesis by teaching the integration of conflicting information from multiple sources. *Cognition and Instruction, 36*(2), 119–138. https://doi.org/10.1080/07370008.2018.1425300

McCrudden, M. T., Magliano, J. P. & Schraw, G. (2011a). The effect of diagrams on online reading processes and memory. *Discourse Processes, 48*(2), 69–92 https://doi.org/10.1080/01638531003694561

McCrudden, M. T., Magliano, J. P., & Schraw, G. (2010). Exploring how relevance instructions affect personal reading intentions, reading goals and text processing: A mixed methods study. *Contemporary Educational Psychology, 35*(4), 229–241. https://doi.org/10.1016/j.cedpsych.2009.12.001

McCrudden, M. T., Magliano, J. P., & Schraw, G. (2011b). Toward an integrated view of relevance in text comprehension. In M. T. McCrudden, J. P. Magliano & G. Schraw (Eds.), *Text relevance and learning from text* (pp. 395–414). IAP Information Age Publishing.

McCrudden, M. T., & Schraw, G. (2007). Relevance and goal-focusing in text processing. *Educational Psychology Review, 19*(2), 113–139. https://doi.org/10.1007/s10648-006-9010-7

McCrudden, M. T., & Schraw, G. (2010). The effects of relevance instructions and verbal ability on text processing. *The Journal of Experimental Education, 78*(1), 96–117. https://doi.org/10.1080/00220970903224529

Moreno, J. D., León, J. A., Kaakinen, J. K., & Hyönä, J. (2020). Relevance Instructions Combined with Elaborative Interrogation Facilitate Strategic Reading: Evidence from Eye Movements. *Psicología Educativa, 27*(1), 51–65. https://doi.org/10.5093/psed2020a20

Perkins, D. N., & Salomon, G. (1992). Transfer of learning. In T. Husén, & T. N. Postlethwaite (Eds.), *The international encyclopedia of education* (pp. 425–441). Oxford: Pergamon.

Reigeluth, C. M., & Stein, F. S. (1983). The elaboration theory of instruction. In C. M. Reigeluth (Ed.), *Instructional design theories and models: An overview of their current status* (pp. 335–382). Hillsdale, NJ: Erlbaum.

Renkl, A. (2014). Learning from worked examples: how to prepare students for meaningful problem solving. In V. A. Benassi, C. E. Overson & C. M. Hakala (Eds.), *Applying science of learning in education: infusing psychological science into the curriculum* (pp. 118–130). Washington: Society for the Teaching of Psychology.

Renkl, A., Mandl, H., & Gruber, H. (1996). Inert knowledge: Analyses and remedies. *Educational Psychologist, 31*(2), 115–121. https://doi.org/10.1207/s15326985ep3102_3

Roelle, J., Lehmkuhl, N., Beyer, M.-U., & Berthold, K. (2015). The role of specificity, targeted learning activities, and prior knowledge for the effects of relevance instructions. *Journal of Educational Psychology, 107*(3), 705–723. https://doi.org/10.1037/edu0000010

Rosenshine, B., & Meister, C. (1992). The use of scaffolds for teaching higher-level cognitive strategies. *Educational Leadership*, *49*(7), 26–33.

Rouet, J. F., & Britt, M. A. (2011). Relevance processes in multiple documents comprehension. In M. T. McCrudden, J. P. Magliano & G. Schraw (Eds.), *Text relevance and learning from text* (pp. 19–52). Charlotte, NC: Information Age Publishing.

Schellenbach-Zell, J., & Neuhaus, D. (2022). Ein Lehretandem für das Praxissemester: Chancen, Risiken, Limitationen. *Die Materialwerkstatt. Zeitschrift für Konzepte und Arbeitsmaterialien für Lehrer*innenbildung und Unterricht*, *4*(1), 98–115. https://doi.org/10.11576/dimawe-5885

Schneider, M. (2012). Knowledge integration. In N. M. Seel (Ed.), *Encyclopedia of the sciences of learning* (pp. 1684–1686). New York: Springer. https://doi.org/10.1007/978-1-4419-1428-6_807

Seel, A. (1997). Von der Unterrichtsplanung zum konkreten Lehrerhandeln - Eine Untersuchung zum Zusammenhang von Planung und Durchführung von Unterricht bei Hauptschullehrerstudentinnen. *Unterrichtswissenschaft, 25*(3), 257–273.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15*(2), 4–14. https://doi.org/10.3102/0013189X015002004

Smith, B. L., Holliday, W. G., & Austin, H. W. (2010). Students' comprehension of science textbooks using a question-based reading strategy. *Journal of Research in Science Teaching, 47(*4), 363-379. https://doi.org/10.1002/tea.20378

Thompson, N. (2010). KR-20. In N. J. Salkind (Ed.), *Encyclopedia of research design* (pp. 668–668). Thousand Oaks: SAGE Publications. https://dx.doi.org/10.4135/9781412961288.n205

Tophinke, D. (2021). Rechtschreiben. In G. Lange & S. Weinhold (Eds.), *Grundlagen der Deutschdidaktik* (pp. 101–127). Baltmannsweiler: Schneider Verlag.

Tynjälä, P., Mason, L., & Lonka, K. (2001). Writing as a learning tool: An introduction. In P. Tynjälä, L. Mason & K. Lonka (Eds.), *Writing as a learning tool: Integrating theory and practice* (pp. 7-22). Dodrecht: Kluwer Academic Publishers. https://doi.org/10.1007/978-94-010-0740-5_2

Wäschle, K., Lehmann, T., Brauch, N., & Nückles, M. (2015). Prompted journal writing supports preservice history teachers in drawing on multiple knowledge domains for designing learning tasks. *Peabody Journal of Education, 90*(4), 546–559. https://doi.org/10.1080/0161956X.2015.1068084

Weinert, F. E., Schrader, F.-W., & Helmke, A. (1990). Educational expertise: Closing the gap between educational research and classroom practice. *School Psychology International, 11*(3), 163–180. https://doi.org/10.1177/0143034390113002

Wiley, J., & Voss, J. F. (1999). Constructing arguments from multiple sources: Tasks that promote understanding and not just memory from text. *Journal of Educational Psychology, 91*(2), 301–311. https://doi.org/10.1037/0022-0663.91.2.301

Zeeb, H., Biwer, F., Brunner, G., Leuders, T., & Renkl, A. (2019). Make it relevant! How prior instructions foster the integration of teacher knowledge. *Instructional Science, 47*(6), 711–739. https://doi.org/10.1007/s11251-019-09497-y

Zeeb, H., Spitzmesser, E., Röddiger, A., Leuders, T., & Renkl, A. (2020). Using relevance instructions to support the integration of teacher knowledge. In T. Lehmann (Ed.), *International perspectives on knowledge integration: Theory, Research, and Good Practice in Pre-service Teacher and Higher Education* (pp. 201–229). Leiden: Brill | Sense. https://doi.org/10.1163/9789004429499_010

Tables

Table 1

Description of textbook excerpts provided as learning material.

Professional knowledge domain	Content description	Source	Frequency of words	Frequency of sentences	Average sentence length (in words)	Readability index (LIX)
CK	Orthography theory: Grammatical principles of spelling, grapheme-phoneme correspondences, other orthographic principles	Tophinke (2021)	1516	80	18.9	58.9
PK	Effective teaching through cognitive activation and metacognitive support	Lipowsky (2020)	1505	57	26.4	73.4
PCK	Spelling awareness and spelling consciousness: prerequisites for spelling-conscious (thought) actions, spelling competence, and their development	Hanisch (2018)	1554	69	22.5	65.7

Note. The texts were slightly modified from their original version for the sake of internal validity (i.e., alignment of typeface and layout, removal of marginalia, standardization of references). Readability indices (LIX) were estimated using the software from Lenhard & Lenhard (2017), which applies the formula of Björnsson (1983). Each LIX represents the sum of the average sentence length of a text and the percentage of long words (more than six letters). As a rule, the following list is used as an estimate of difficulty, based on the average values of various text genres: Below 40: Children's and Youth Literature; 40 to 50: Fiction; 50 to 60: Non-fiction; Above 60: Academic Literature. As a limitation, it should be noted that the index captures only a limited set of factors that determine a text's difficulty. For instance, a sentence might be long but still easy to understand if it's just a list of bullet points. Also, surface features such as word frequency are not considered. On the other hand, a text may appear simple on the surface but become complex because it introduces unfamiliar content. In summary, the LIX is useful for an approximate estimation of text complexity, but does not adequately represent all complexity features.

Table 2

Sample items and internal consistency coefficients of the domain-specific knowledge tests.

Test	Sample items	Pre-test KR20	Post-test KR20
CK-test	*Closed question sample (multiple choice with multiple correct answers):* What principles can be clearly attributed to a grammaticalization of spelling? a) The morphological principle [correct] b) The aesthetic principle [incorrect] c) The etymological principle [incorrect] d) The syntactic principle [correct] *Open question sample:* Explain the basic assumption on which the phoneme-grapheme correspondence concept is based.	.63	.69
PK-test	*Closed question sample (single choice with a single correct answer):* A teacher wants to improve her students' metacognitive skills. She therefore decides to integrate the following measure into her lessons: a) Phases of self-assessment [correct]. b) Phases of critical discussion of the learning content in groups [incorrect] c) Phases of individual practice [incorrect] d) Phases of prior knowledge activation [incorrect] *Open question sample:* How does the construct of cognitive activation relate to the constructivist learning theories of Vygotsky and Piaget?	.57	.63
PCK-test	*Closed question sample (multiple choice with multiple correct answers):* Which adjectives describe the concept of spelling awareness? a) contextualized [incorrect] b) verbalizable [correct] c) automatic [incorrect] d) volitional [correct] *Open question sample:* What role does spelling intuition play in the development of reflective and cognitive spelling awareness?	.59	.65

Note. Annotations in square brackets. KR20 is an estimate of the internal consistency of scales with dichotomous items and is based on the Kuder-Richardson Formula 20.

Table 3

Coding scheme for the assessment of students' essays.

Measurement dimension	Coding category	Description	Example
Separative learning	Borrowing	Student restates information from a text without changing the meaning expressed in the domain-specific source, e.g., via a citation or paraphrase.	"The subject-matter related level of the didactic communication and class discussions plays is particularly important for a cognitively activating lesson." [borrowing from PK text]
	Elaboration within domains	Student elaborates on a domain-specific idea using further information from the same text and/or from prior knowledge without making novel connections across domains.	"But even within orthographic theory, the (non-) grammatical principles of spelling are not clearly defined, as shown, for example, by the morphological and sematic principles." [separative elaboration within the CK domain]
Integrative learning	Elaboration across domains	Student combines information from one domain-specific text with information from one or both other texts which represent different knowledge domains. The student potentially integrates further information from prior knowledge (i.e., information on the same topic which is not presented in any of the texts).	"A strategy for the cognitive activation of learners in spelling lessons could be to discuss different spellings of words that conform to a rule but are still incorrect - first in partner work, then in the plenary." [integrative elaboration of PK- and PCK-specific information]
	Switches	Frequency of switches between domains	"Phoneme-grapheme correspondences are central to German spelling, but contain many cases of ambiguous assignments. [Switch from CK to PCK] Therefore, focusing on rule-based spelling skills in the classroom is not enough."
Addition	Addition of opinion, pre-knowledge and metacognitions	Student includes relevant information from prior knowledge (without linking it to other statements) or expresses personal opinion or metacognitive thoughts.	"Authentic writing tasks that are relevant to the pupils' life world have a motivating effect and are therefore conducive to learning." [pre-knowledge] "Pupils are mostly bored when it comes to proper writing." [personal opinion] "For me, it is more difficult to write about connections between the subject-matter text [CK] and the other two texts [PK, PCK] than just between the other two." [metacognitive information]

Note. English translations of participants' original wording in their essay. Additional coding information added in squared brackets.

Table 4

Categories, examples, and ICCs for coding participants' responses to the integrated knowledge application test.

Category	Example	ICC
CK application	"Listening task" [positive feature]: "The words read out by the teacher in the listening task have been selected correctly, i.e., they are suitable to show that morphologically related forms are spelled similarly (e.g. <Baum> / <Bäume>)."	.94
PK application	"Class arrangement" [feature for improvement]: "Overall, the lesson involves too much frontal teaching; by including group work phases the cognitive activation of the pupils could probably be increased, since the students would then engage more intensively in exchanging ideas."	.90
PCK application	"Lessons learned" [positive feature]: "Having pupils formulate a perceived linguistic regularity, in this case the morphological principle, as a rule contributes to their explicit spelling knowledge (spelling awareness)."	.87
Incorrect knowledge application	"Exposing the children" [feature for improvement]: "I would recommend not to ask which children want to read out their written stories (homework), because this will embarrass some."	.82
Pre-knowledge application	"Internal differentiation" [improvable feature]: "The worksheet for individual work is the same for everyone. However, it should be individualized according to the different learning requirements to be assumed, since the task could be otherwise over- or underwhelming."	.83

Note. English translations of participants' original wording. Additional information added in squared brackets.

Table 5

Means and standard deviations of study and teaching experience, time on task, idea units, lesson plan features, and domain-specific prior knowledge.

		Experimental condition			
		CG	RI	GQ	
	Measure	*M (SD)*	*M (SD)*	*M (SD)*	ANOVA
Study experience	Semester	7.78 (1.16)	7.50 (.962)	7.71 (.937)	*F*(2,80)=.563, *p*=.572
Practical teaching experience	Internship duration in months	6.30 (4.31)	6.48 (4.00)	6.40 (3.73)	*F*(2,80)=.014, *p*=.986
	Lessons planned and held	19.53 (18.14)	20.98 (18.84)	16.34 (11.75)	*Welch's F*(2,50.29)=.715, *p*=.494
Time on task	Time spent on reading/writing in minutes	58.26 (7.71)	58.94 (7.57)	57.73 (7.92)	*F*(2,80)=.175, *p*=.840
Idea units included in the essay	Sum of idea units	19.07 (5.01)	18.32 (4.29)	19.64 (5.39)	*F*(2,80)=.509, *p*=.603
Features identified in the lesson plan to be evaluated	Sum of features	5.70 (1.20)	5.43 (.920)	5.29 (1.21)	*F*(2,80)=.987, *p*=.377
Prior knowledge	CK pre-test	3.37 (1.57)	3.11 (1.45)	3.71 (1.56)	*F*(2,80)=1.11, *p*=.334
	PK pre-test	2.78 (1.67)	2.39 (1.79)	2.36 (1.45)	*F*(2,80)=.551, *p*=.578
	PCK pre-test	2.70 (1.56)	2.89 (1.34)	3.00 (1.28)	*F*(2,80)=.316, *p*=.730

Note. CG = control group, RI = relevance instruction, GQ = guiding questions.

Table 6

Means and standard deviations of the domain-specific knowledge gains

		Experimental condition		
		CG	RI	GQ
	Measure	*M (SD)*	*M (SD)*	*M (SD)*
Domain-specific knowledge acquisition	ΔCK	3.52 (1.19)	3.75 (1.01)	3.57 (.997)
	ΔPK	4.74 (1.72)	4.64 (1.79)	4.46 (1.71)
	ΔPCK	3.74 (1.20)	4.00 (1.52)	3.89 (1.52)

Note. CG = control group, RI = relevance instruction, GQ = guiding questions. The Δ-values reflect the difference between the pre- and post-knowledge test.
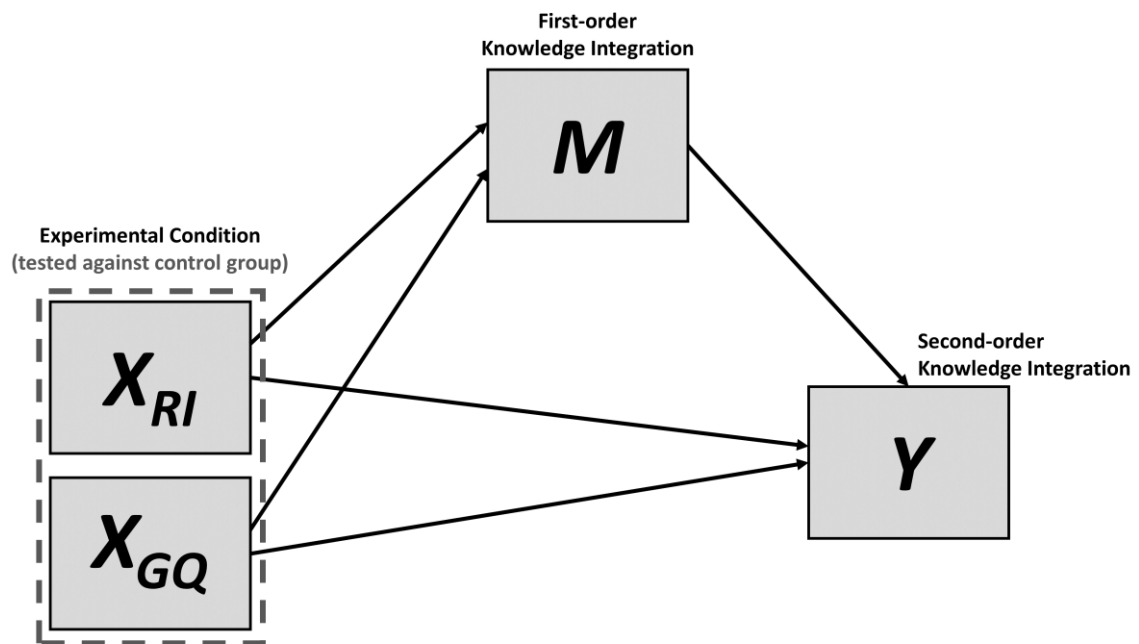
Figures

Figure 1. The conceptual model of the mediation hypothesis (H3). Abbreviations of experimental conditions: RI = Relevance Instruction, GQ = Guiding Questions.
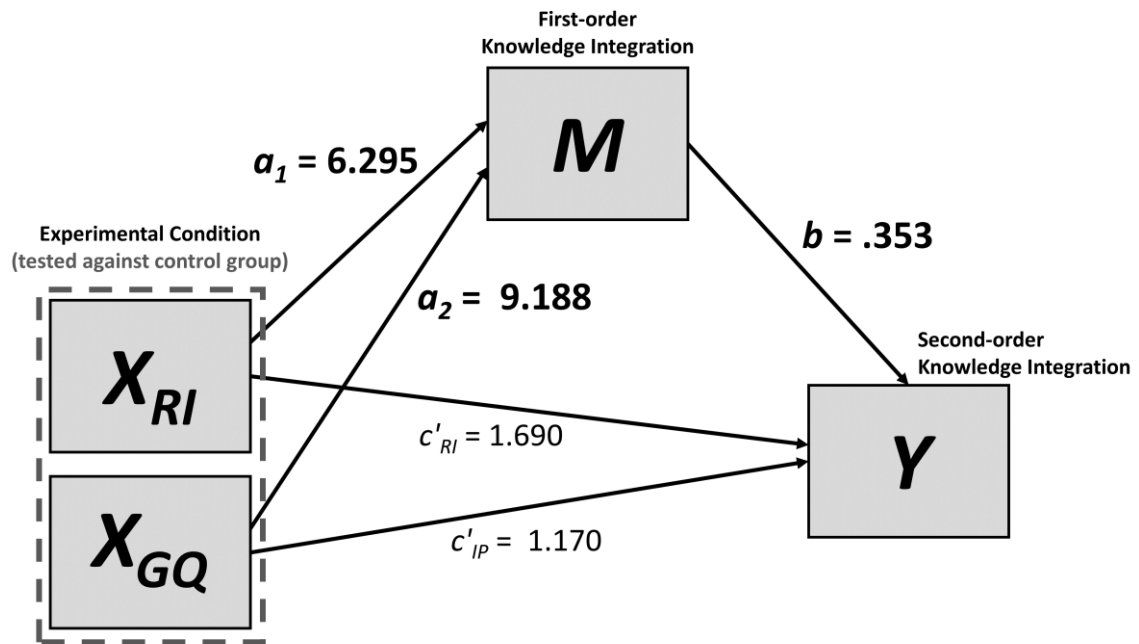
Figure 2. The statistical model of the mediation hypothesis H3 (significant coefficients in bold). Abbreviations of experimental conditions: RI = Relevance Instruction, GQ = Guiding Questions.