



Fostering integrated mental models of different professional knowledge domains: instructional approaches and model-based analyses

Thomas Lehmann¹ · Pablo Pirnay-Dummer² · Florian Schmidt-Borcherding¹

© The Author(s) 2019, corrected publication 2021

Abstract

Recent research on expert teachers suggests that an integrated understanding across the core domains of teachers' knowledge is crucial for their professional competence. However, in initial teacher education pre-service teachers seem to struggle with the integration of knowledge represented in multiple domain-specific sources into a coherent structure (e.g., textbooks that focus either on content knowledge, on content-specific pedagogical knowledge, or on general pedagogical knowledge). The purpose of this study was to investigate the effectiveness of writing tasks (unspecific vs. argumentative) and prompts (i.e., focus questions) on pre-service teachers' construction of a mental model that interrelates information from multiple domain-specific documents. Data of ninety-two pre-service teachers, who participated in a laboratory experiment where they read three domain-specific textbook excerpts and wrote essays for global comprehension, were analyzed using automated structural and semantic measures. In line with prior research, results indicated that prompts supported pre-service teachers in integrating domain-specific knowledge from multiple documents in their mental models. However, the automated structural and semantic measures did not support previous findings on the efficacy of argument tasks for knowledge integration. The findings and limitations are discussed, and conclusions are drawn for future research and for integrative learning environments in pre-service teacher education.

Keywords Mental models · Knowledge integration · Assessment · Knowledge representation · Pre-service teacher education · Teacher knowledge

✉ Thomas Lehmann
tlehmann@uni-bremen.de

Pablo Pirnay-Dummer
pablo@pirnay-dummer.de

Florian Schmidt-Borcherding
florian.schmidt-borcherding@uni-bremen.de

¹ Research Unit: Learning, Instruction, and Educational Psychology, Faculty of Pedagogical and Educational Sciences, University of Bremen, Universitäts-Boulevard 11/13, 28359 Bremen, Germany

² Department of Educational Psychology, Faculty of Pedagogy, Martin-Luther-University Halle-Wittenberg, Franckeplatz 1, Haus 5, 06099 Halle, Germany

Introduction

The academic phase of initial teacher education typically initiates the development and improvement of pre-service teachers' content knowledge (CK), general pedagogical knowledge (PK), and pedagogical content knowledge (PCK). This educational practice is in line with Shulman's (1986) classic taxonomy of teacher knowledge and recent models of teachers' professional competence (e.g., Baumert and Kunter 2013).

According to these conceptualizations, CK is the basic requirement for teaching the subject-matter content to pupils. It incorporates (a) the subject matter knowledge to be taught, (b) a deeper understanding of the subject matter knowledge to be taught, and (c) an awareness and understanding of the relationships between the subject's topics included in the curriculum (Ball et al. 2008; Krauss et al. 2008). PK, on the other hand, is usually defined as the general pedagogical knowledge that is needed for professional teaching independent of the subject to be taught. It involves "broad principles and strategies of classroom management and organization that appear to transcend subject matter" (Shulman 1987, p. 8), which enable teachers to optimize learning situations in general (see Voss et al. 2011 for a short overview). Finally, PCK "is different from content knowledge, on the one hand, because of the focus on the communication between teacher and student, and from general pedagogical knowledge, on the other, because of the direct relationship with subject matter" (Verloop et al. 2001, p. 449). It incorporates knowledge about students' typical conceptions and learning difficulties, and knowledge about effective representations and teaching strategies, both with regard to particular topics of the subject matter or the content domain in general. However, initial teacher education is usually fragmented and the academic teachings of these "knowledge domains" (Baumert and Kunter 2013, p. 291) are rarely systematically linked beyond their conceptual boundaries (Darling-Hammond 2006). Accordingly, learning material such as text documents (e.g., textbooks, journal articles, lecture scripts) generally focuses either CK, PK, or PCK in detail, and does not provide an integrated perspective as regards teaching. Yet, in order to not only gain CK, PK, and PCK but also improve the quality of their knowledge's mental representation, pre-service teachers have to initiate and regulate the integration of their knowledge in a self-regulated manner.

Knowledge integration is characterized as a "dynamic process of linking, connecting, distinguishing, organizing, and structuring [...] patterns, templates, views, ideas, theories, and visualizations" (Linn 2000, p. 783). But the self-directed integration of 'models' from different teacher knowledge domains is a very demanding task for pre-service teachers, especially without any instructional stimulation or support (Wäschle et al. 2015), or various practical experiences (Hashweh 2005). As a result, student teachers' professional knowledge is hardly integrated when they complete their academic phase of initial teacher education (Ball 2000; Darling-Hammond 2006). They possess only few links that bridge the gaps between knowledge structures and entities that are compartmentalized in separate memory parts (Renkl et al. 1996). Hence, they have hard times taking multiple intelligent perspectives (Weinert et al. 1990) in, for example, lesson planning (Seel 1997), modifying existing learning material from textbooks for their lesson (Hashweh 1987), or designing learning assignments (Wäschle et al. 2015), which all are considered as key tasks for professional teaching (Baumert and Kunter 2013).

Since teaching experts do not only possess more professional knowledge (i.e., CK, PK, and PCK) than novices but especially differ by the extent of how much cross-references link initially unconnected knowledge structures into a common mental model (e.g.,

Berliner 2001; Bromme 2014; Livingston and Borko 1990), the development of purposeful learning opportunities is required to foster not only pre-service teachers' acquisition of professional knowledge but also their integration of knowledge. Few studies recently addressed the question of enhancing pre-service teachers' integration of knowledge that stems from different professional domains (Harr et al. 2014, 2015; Janssen and Lazonder 2016; Lee and Turner 2017; Lehmann et al. 2019; Wäschle et al. 2015). Some studies investigate writing tasks (Lehmann et al. 2019) and prompted journal writing (Wäschle et al. 2015) since writing itself has been often recognized as a tool for learning in general, and for integrative knowledge-transforming in particular (Scardamalia and Bereiter 1991; Tynjälä et al. 2001; Wiley and Voss 1999). Moreover, *writing-to-learn* involves externalization, which enables learners to not only understand but also communicate their understanding of the learning content by using adequate sign and symbol systems (Le Ny 1993), for example, via written formulations such as essays or learning journals. Externalized models then can be used by researchers to gain insight into the learners' internal model (Pirnay-Dummer 2015; Seel 1991). Yet, findings on pre-service teachers' integration of different domain-specific knowledge structures are to some extent ambiguous depending on the method of assessment. As a result, there is not only a need for replication and further studies but also for new ways of assessment.

Taking the aforementioned as a starting point, the present study aims at examining (a) an argument writing task in comparison to an unspecific writing task for global comprehension, and (b) task-supplemental prompts with regard to their effect on knowledge integration. The analysis are preformed using automated structural and semantic measures, which found on the theory of mental models (Johnson-Laird 1983; Seel 1991), and have—to the best of our knowledge—not been implemented in studies on pre-service teachers' knowledge integration across multiple domains to date.

Mental models in (multiple) text comprehension research

Kintsch (1974) took a first step to investigate text comprehension by focusing on the higher order processes of reading as compared to the processes of basal text processing such as the recognition of words. Kintsch's model emphasized that the meaning of a text is represented by several mentally constructed and interconnected semantic units, so-called propositions. However, the model was criticized since the propositional representations seemed to be more representative for the text structure as compared to the structure of memory for the text (McNamara et al. 1991). As a result, enhancements of the model such as the construction-integration (CI) model (van Dijk and Kintsch 1983) implied a more holistic understanding of reading comprehension and prior knowledge was considered as an additional key factor.

According to the CI model, readers develop their understanding of a text in a two-phased cycle. In the construction phase, the semantic content of a text is extracted by identifying propositions which activate a network of related, both relevant and irrelevant concepts from prior knowledge. In the integration phase, the more or less "chaotic" network of activated propositions and concepts serves as an input. In a constraint-satisfaction process, propositions that are perceived as helpful (e.g., due to consistence with prior knowledge) are kept active whereas unbecoming propositions are discarded. Moreover, the CI model suggests a distinction between two types of representations. Readers develop (1) a text base, that is, a mental representation of the semantic content which is constructed from propositions; and they construct (2) a situation model (or *mental*

model), which represents the reader's interpretation of what is described in the text rather than representing the text itself. It is the mental model which allows the reader to understand things that are not explicitly stated in the text (van Dijk and Kintsch 1983; Johnson-Laird 1983).

Although the CI model had a major impact on researchers' understanding of students' comprehension of a text, the model is rather limited in explaining text-based learning. The reason for this lays in the fact that according research focused on text comprehension processes when reading a single, relatively simple text with clear reading goals. More complex processes of understanding text sources thus remained unnoticed. In order to accommodate the complexity of dealing with multiple information sources Perfetti et al. (1999) introduced the "documents model" framework. This framework suggests that an expert reader of multiple documents gives not only credit to what is actually written in the documents (i.e., by constructing multiple but isolated text bases), and what this means (i.e., by constructing an integrated mental model of the documents' contents); he/she also considers various document characteristics, such as author, setting, form, rhetorical goals (i.e., intent, audience), and content (i.e., thesis) to in- and exclude specific source information.

More recently, Rouet and Britt (2011) proposed the Multiple-Document Task-Based Relevance Assessment and Content Extraction (MD-TRACE) model. This model pays special attention to the task model, that is, "a subjective representation of the goal to be achieved and the means available to achieve it" (Rouet et al. 2017, p. 207) in a multiple text comprehension setting. McCrudden and Schraw (2007, 2010) examined relevance instructions (i.e., *prompts* in the form of focus questions or study objectives that serve as "strategy activators"; Reigeluth and Stein 1983) with regard to stimulating students' provision of explanations, taking of specific positions while reading, and studying the texts for a particular goal such as passing an exam. Their findings provide empirical support for the MD-TRACE model in that readers' cognitive processing and construction of an integrated mental model is strongly influenced by the reader's task model. Hence, the notion that the task model "drives subsequent processes involved in the search, evaluation and integration of information" (Rouet and Britt 2011, p. 31) is widely accepted (see also Kopp 2013; List and Alexander 2017).

Aside from prompts as a task-supplemental instructional scaffold to stimulate specific learning strategies (Lehmann et al. 2014; Bannert, 2009; Reigeluth and Stein 1983), researchers interested in students' cognitive processes of multiple document comprehension have also investigated different types of *writing tasks* with regard to their influence on integrated mental models of multiple texts. For example, Bråten and Strømsø (2009) compared three types of writing task instructions (i.e., argument, summary, global understanding). They found that students who received the argument or summary task performed better in an inference verification task across the learning sources' contents after controlling for gender, age, word decoding, and prior knowledge than students who received the writing task for global understanding which is less specific in its generic sense.

Wiley and Voss (1996, 1999) also examined the effect of different writing tasks on students' multiple text comprehension. They compared the effects of instructing students to write either a narrative, a summary, a history, an explanation, or an argument. Their results showed that argument writing tasks provoke a higher degree of transformation, integration, and causality in students' essays compared to the other writing tasks. Moreover, the studies provided evidence for a positive influence of argument tasks on the development of a deeper and more integrated understanding of the text documents' contents. Wiley and Voss (1999) attribute this effect to a task-induced change in students' strategic learning behavior from a rather passive reproduction (i.e., knowledge-telling) to an active-constructive

transformation (i.e., knowledge-transforming) of text information (Scardamalia and Bereiter 1991).

Recent research (Lehmann et al. 2019; Wäschle et al. 2015) adopted these findings to pre-service teachers' learning from multiple documents. In two experimental studies, the researchers provided three documents (e.g., journal articles, textbook excerpts) as learning material. Each document was a representative source of one conceptually distinct teacher knowledge domain (i.e., CK, PK, and PCK; Baumert and Kunter 2013). The study of Wäschle et al. (2015) focused the influence of *prompts* on the use of non-integrative and integrative learning strategies in journal writing. For this purpose, fifty-two pre-service history teachers, who already had experience in learning journal writing, were asked to read a journal article about a historical event, a textbook excerpt about the specifics of teaching history, and an educational-psychological textbook excerpt about cognitive learning processes. Then they were instructed to write a learning journal entry about the contents while having all text sources at hand. One of the two experimental groups additionally received prompts in terms of four focus questions, whereas the other did not. It was found that students of the prompted experimental condition engaged more frequent in learning strategies that involve knowledge integration than students in the unprompted condition. Furthermore, it was found that the application of integration strategies predicted pre-service teachers' performance in a learning task evaluation and a problem solving test.

In another study, Lehmann et al. (2019) delivered evidence for the efficacy of both argument tasks and task-supplemental prompts in enhancing pre-service mathematics teachers' construction of integrated mental models of multiple documents. However, the findings varied depending on the applied assessment methods. A category-driven content analysis of the participants' externalized mental models (i.e., the essays as the individual solutions to the writing tasks) indicated that the argument task and the prompts independently influenced students' integrative knowledge-transformation in terms of constructing more integrative elaborations and making more switches between the domain-specific learning sources. Students who worked on the unspecific writing task for global comprehension and who did not receive task-supplemental prompts depicted significantly more often domain-specific ideas that were taken directly or paraphrased from one of the documents without changing or expanding the meaning of the source text. Additionally, after controlling for the influence of participants' interest/enjoyment, their perceived competence, and their self-reported effort it was found that their ability to make intertextual inferences that connect CK, PK, and PCK learning contents was positively affected by prompts that complemented the unspecific writing task for global comprehension. However, contrary to the results, which were based on the analysis of participants' externalized models, results of this intertextual inference verification task did not indicate a significant effect of the argument task on students' multiple text comprehension.

Taking the findings of Wäschle et al. (2015) and Lehmann et al. (2019) together stimulates a demand for alternative methodological approaches to analyze knowledge integration across conceptually distinct domains such as pre-service teachers' integration of CK, PK, and PCK. As will be shown in the following, model-based assessment technology and tools could meet this demands, and thus, complement prior analyses.

Research on the assessment of mental models

The theoretical construct of mental models as discussed above in the context of (multiple) text comprehension allows cognitive and educational researchers to explain a

humans' learning (e.g., in terms of the development of an understanding of various contents presented in multiple documents), as well as their expertise, reasoning, and problem solving (Lehmann and Pirnay-Dummer 2014; Seel 1991). However, the theory of mental models suggests that knowledge and its mental representation makes it impossible to measure mental models directly (Ifenthaler 2010; Seel 1991). Hence, many methodologies that were developed to provide time economic (compared to existing assessment and scoring procedures), valid and reliable assessment systems and automated computer-based diagnostic tools demand the *externalization* of learners' mental models (Ifenthaler and Pirnay-Dummer 2014).

As mentioned earlier, learners can externalize their mental model by way of numerous communication formats using adequate sign and symbol systems (Le Ny 1993). In accordance with semiotics, which differentiates between the use of signs as an index, as an icon or as a symbol (Seel 1999), the cognitive-psychological perspective of Bruner (1964) leads to a differentiation of an enactive, an iconic, and a symbolic format of representation. Depending on the task and the specific requirements of the situation learners can either recall an appropriate form of representation or transform a mental model into an appropriate form of representation, for example, into drawings (Chi et al. 1994; Seufert et al. 2007), causal diagrams (Spector et al. 2005), concept maps (Johnson et al. 2009), or natural language statements and written formulations (Pirnay-Dummer et al. 2010; Pirnay-Dummer 2014; Johnson-Laird 1983). Regardless of the form and modality, an externalized model is always a *re-representation* (Ifenthaler 2010) because it represents what is already a representation, that is, the internal representation (mental model) of conceptual and causal interrelations of "objects, states of affairs, sequences of events, the way the world is, and social and psychological actions" (Johnson-Laird 1983, p. 397).

A related point to consider is that externalization by itself can lead to a modification of a learner's current mental model, and hence, improve understanding (e.g., writing-to-learn: Tynjälä et al. 2001; generating pictures: Seufert et al. 2007). Moreover, it is important to note that any externalization of a mental model also involves the learner's interpretation (Ifenthaler 2010; Ifenthaler and Pirnay-Dummer 2014). More specifically, in multiple document learning settings re-representations involve the learners' interpretation of the instruction that demands externalization (i.e., the task model), and an interpretation of the learning context (Rouet et al. 2017). As a result, re-representations, especially those in form of natural language expressions, not only re-represent a learner's knowledge structure and semantics but also their external reconstruction plus the person's rhetoric and pragmatics (Halford et al. 2010; Leech 1983; van Dijk 1977). This theoretical perspective on natural language re-representations is also supported by Scardamalia and Bereiter's (1987) psychology of written composition by making a distinction between two interacting problem spaces, that is, the content problem space (which includes structure and semantics) and the rhetorical problem space (which includes thoughts about how to adequately represent an internal model). In regards to assessment and analysis, interpretation is also carried out by the methodology applied (according to certain pre-defined rules) and by researchers, who interpret the results. "This mixture [of interpretations] and the complexity of the [mental model] construct both make it specifically difficult to trace the steps and bits of knowledge" (Ifenthaler and Pirnay-Dummer 2014, p. 291). Still, several tools have been successfully developed for the valid and reliable assessment and analysis of mental models or rather of their re-representations, for example, ALA-Reader (Clariana 2010), DEEP (Spector and Koszalka 2004), jMAP (Shute et al. 2010), HIMATT (Pirnay-Dummer et al. 2010), and T-MITOCAR (Pirnay-Dummer and Ifenthaler 2010; Pirnay-Dummer 2015).

T-MITOCAR (Text-Model Inspection Trace of Concepts and Relations) has a background in mental model theory (Johnson-Laird 1983; Seel 1991), association psychology (Davis 1990; Lewin 1922; McKoon and Ratcliff 1992; McNamara 1992, 1994; Ratcliff and McKoon 1994; Stachowiak 1979), and linguistics (Frazier 1999; Pollio 1966; Russel and Jenkins 1954). The tool uses natural language expressions as input data, for example, learners' argumentative essays or other written formulations. It relies on the dependence of syntax and semantics within natural language and uses the associative features of text as a methodological heuristic to represent knowledge from natural language. *T-MITOCAR* allows to analyze written formulations very quickly without loosening the associative strength of natural language and works with a comparable small amount of text (350 words+), too. More specifically, the tool measures structural and semantic similarities of verbally externalized mental models compared to a reference (expert) model. Table 1 provides descriptions for each measure included in the *T-MITOCAR* assessment tool.

In a study on evaluating model-based analysis tools in the mathematics domain, Gogus (2012, 2013) found that *T-MITOCAR*'s *GAMMA* "does provide a basis for determining a relative level of expertise and may help in identifying gaps in a novice representation" (Gogus, 2013, p. 188). Hence, *GAMMA* is of particular importance for the present study as the focus is on the construction of well-connected (i.e., integrated) mental models versus fragmented models that include separate, unconnected knowledge entities. In regard to the semantic dimension of mental modelling, the *BSM* measure is particularly important for this study. While *CONC* looks simply for matching concepts, *PROP* looks simply for matching edges (i.e., links, relations) between two graphs. Hence, a propositional match already demands two matching concepts. "But propositions obviously do not automatically match only because the concepts do. Sometimes this dependency is hard to interpret from the first two semantic measures alone, especially when individual comparisons are aggregated otherwise (e.g., within group means)" (Kopainsky et al. 2010, p. 21). The *BSM* accounts for this dependency, and can therefore be preferred in domains with complex (causal) relationships (cf. Barry et al. 2014).

Present study and hypotheses

Until to date, there are no studies that use *T-MITOCAR* to examine learners' externalized mental models with regard to knowledge integration in the field of pre-service teacher education, which requires to integrate knowledge from multiple domains. Yet, prior research suggests that *T-MITOCAR* is a promising tool for the analysis of learners' interrelating/merging of information from different knowledge domains and their construction of an integrated mental model of multiple documents.

In this article, we present a study which sought to investigate the efficacy of argument tasks and prompts on the construction of integrated mental models of different professional knowledge domains by using specific measures of *T-MITOCAR* as an automated model-based analysis tool. Considering the recent empirical findings of Lehmann et al. (2019), we expected a positive effect of task-supplemental integration prompts (Hypothesis H1a) and argument writing tasks (Hypothesis H1b) and on pre-service teachers' construction of an integrated model in terms of structural connectedness (i.e., the density of vertices in their written knowledge re-representations). Furthermore, we assumed that the models of students who do not receive prompts (Hypothesis H2a) and receive an unspecific writing task (Hypothesis H2b) depict a higher semantic-propositional similarity (i.e., the quotient of

Table 1 Description of the structural and semantic measures of T-MITOCAR (cf. Iffenthaler 2011; Pirnay-Dummer et al. 2010)

Category of measurement	Measure	Description
Structural	Gamma-matching (GAMMA)	The gamma or density of vertices (Pirnay-Dummer and Iffenthaler 2010) describes the quotient of terms per vertex within a graph. Since both graphs which connect every term with each other term (everything with everything) and graphs which only connect pairs of terms can be considered weak models, a medium density is expected for most good working models
	Surface-matching (SUR)	The surface matching (Iffenthaler 2010) compares the number of vertices within two graphs. It is a simple and easy way to calculate values for surface complexity
	Graphical-matching (GRA)	The graphical matching (Iffenthaler 2010) compares the diameters of the spanning trees of the graphs, which is an indicator for the range of conceptual knowledge. It corresponds to structural matching as it is also a measure for structural complexity only
	Structural-matching (STRU)	The structural matching (Pirnay-Dummer and Iffenthaler 2010) compares the complete structures of two graphs without regard to their content. This measure is necessary for all hypotheses which make assumptions about general features of structure (e.g., assumptions which state that expert knowledge is structured differently from novice knowledge)
	Balanced semantic-matching (BSM)	The balanced propositional matching (Pirnay-Dummer and Iffenthaler 2010) is the quotient of propositional matching and concept matching (see below). Especially when both indices are being interpreted, balanced propositional matching should be preferred over propositional matching
Semantic	Concept-matching (CONC)	Concept matching (Pirnay-Dummer and Iffenthaler 2010) compares the sets of concepts (vertices) within a graph to determine the use of terms. This measure is especially important for different groups which operate in the same domain (e.g. use the same textbook). It determines differences in language use between the models
	Propositional-matching (PROP)	The propositional matching (Iffenthaler 2010) value compares only fully identical propositions between two graphs. It is a measure for quantifying semantic similarity between two graphs

fully identical propositions and concepts) with the reference model, because they process information less transformative.

Overall, this follows a theoretical model that constitutes the complex multiple-documents literacy (in terms of the application of specific cognitive knowledge integration processes such as generating and providing integrative elaborations, and making switches between sources to intertwine domain-specific information and knowledge entities) as a function of structure, semantics, reconstruction, rhetoric, and pragmatics (Halford et al. 2010; Johnson-Laird 1983; Leech 1983; Scardamalia and Bereiter 1987, 1991; van Dijk 1977).

Method

Sample and design

Overall, one-hundred pre-service primary school mathematics teachers from a Northern German university participated in this study. Seven participants had to be excluded due to missing data. One participant had to be excluded since the data was provided in note form as opposed to a written formulation. The remaining ninety-two participants had an average age of $M=24.7$ years ($SD=4.16$), and studied for $M=7.01$ semesters ($SD=1.90$). Most were female (86%), which is typical for pre-service primary school teachers in Germany. All participants were native speakers of German. They took part in the experiment for course credit. To address our research question on the effectiveness of argument tasks and integration prompts, we used a two-factorial between-subject design with four experimental conditions (see Table 2), that is, an unspecific writing task condition (UWT; $n=22$), an unspecific writing task with prompts condition (UWT+P; $n=24$), an argument writing task condition (AWT; $n=24$), or an argument writing task with prompts condition (AWT+P; $n=22$).

Materials

Writing tasks

Writing tasks, either unspecific or argumentative and both for global comprehension in their generic sense, were given on a front page together with the text documents. We did not want to leave the overriding task objective unclear, that is, to not only understand the domain-specific contents but *integrate* them into a more complex, coherent interrelated

Table 2 Depiction of the 2×2 -factorial design of the experimental study

Factor 1	Factor 2	
	Without prompts	With prompts
Unspecific writing task	UWT: $n_1=22$	UWT+P: $n_2=24$
Argumentative writing task	AWT: $n_3=24$	AWT+P: $n_4=22$

UWT unspecific writing task without prompts, *UWT+P* unspecific writing task with prompts, *AWT* argumentative writing task without prompts, *AWT+P* argumentative writing task with prompts

model. Thus, a short written introduction asked participants in any experimental condition to “read the texts carefully and try to understand them as a whole”. In order to enable the unspecific writing task to provide as much freedom as possible for the learning phase, the exact instruction was “write down your thoughts” without asking for any specific text genre. Hence, participants in the UWT condition were put in a writing-to-learn situation in which they were relatively free in their decisions about the content problem space (as framed by the learning material) and the rhetoric problem space (Scardamalia and Bereiter 1991; Tynjälä et al. 2001). The argument task, on the other hand, instructed to “write an argument about how the contents of the various documents are related”. With regard to the content problem space, participants in the argument conditions were as free as participants in the unspecific writing task conditions since their content problem space was also solely framed by the learning material and their own decisions (however, please note that against the background of pre-service teacher education and the learning content, the argument task was considered as provoking a linguistic act in which the relation between CK-, PK, and PCK-specific information needs to be clarified by making it largely undisputed). In contrast, their rhetoric problem space was more predetermined in that they had to write an *argumentative* essay, which in turn should lead to an advantage over an unspecific writing task regarding the integration of source specific information (Bråten and Strømsø 2009; Wiley and Voss 1996, 1999).

Integration prompts

In order to apply prompts as “strategy activators” (Reigeluth & Stein, 1983) for intertextual knowledge integration, we developed five prompts in the form of focus questions under consideration of previous studies that focus on multiple document comprehension (e.g., McCrudden and Schraw 2007, 2010) and the cognitive integration of conceptually distinct knowledge domains in teacher education (Wäschle et al. 2015). The focus questions were worded in a way that aimed at stimulating pre-service teachers’ integration of information from different sources on CK, PK, and PCK (e.g., “Can you identify information in the texts that can be linked to conclusions for the design of lessons which are reasonable from multiple perspectives?”, and “By which examples can you interrelate the texts with regard to learning tasks for your future teaching?”; Lehmann et al. 2019). The prompts were task-supplemental in that they complemented the writing task in order to assist the construction of an adequate task model (which subsequently enables goal oriented search, evaluation, and integration of information from multiple sources; Rouet and Britt 2011). Both the task sheet and the prompts remained with the learners so they could be reread throughout their task performance phase.

Learning materials

Three textbook excerpts, each pertaining to one of the core domains of participants’ professional knowledge, were provided as learning materials (see Table 3). The excerpts stem from textbooks which are frequently used in German pre-service teacher education. As elaborated in Lehmann et al. (2019), the text documents were chosen due to the following characteristics: They represented authentic material that is used in academic teacher education courses. The students are taught to use scientifically reliable sources in their studies and they usually prefer suchlike textbook documents over more scientific publications such as peer-reviewed journal articles. Each document allowed to make connections between

Table 3 Description of excerpts taken from teacher education textbooks and used as learning material

Professional knowledge domain	Content description	Source
CK	Logics and mathematical proving	Grieser (2015)
PK	Learning relevant cognitive and metacognitive processes and some consequences for classroom practice	Renkl (2015)
PCK	Pupils' individual and social difficulties in mathematical reasoning and proving and a process model for teaching mathematical reasoning	Brunner (2014)

The text documents were comparable in readability (see Lehmann et al. 2019) and in length ($M_{\text{words}} = 1946.67$; $SD_{\text{words}} = 238.02$; $M_{\text{sentences}} = 120.67$; $SD_{\text{sentences}} = 5.51$; each on three pages). Together, the three documents provided the content to be learned and integrated. Further variables that have been found to influence multiple document comprehension and knowledge integration such as author expertise, document type, and currentness were maintained constant

information that is presented in one or both of the other documents. Hence, the learning material set was ecologically valid. More information on the documents and their contents to be learned and integrated into a common mental model is displayed in Table 3 (see Lehmann et al. 2019 for further details).

Procedure

In the laboratory experiment, participants were welcomed and asked to fill out a short survey on some personal details (i.e., age, gender, first language etc.), first. Next, they were shortly introduced to the process of the study. Then, they received the writing task and the integration prompts (i.e., five task-supplemental focus questions) according to their experimental condition (see Table 2), and the three textbook excerpts that exemplarily represented the core domains of teachers' professional knowledge and served as domain-specific learning material (see Table 3). The students then worked on their task solutions in a self-regulated manner with all documents at hand throughout the entire process. They were provided as much time as needed to complete the writing task. In average, participants chose to work about 75 min until they handed the textbook excerpts and their essay to the experimenter. As reported in Lehmann et al. (2019), there were no pre-experimental differences in CK, PK, and PCK between the experimental groups and no significant differences regarding the time spent on task.

Data analysis

In the present study the analysis of participants' essays was realized with T-MITOCAR, a tool which utilizes natural language expressions as input data for a *structural* and *semantic* analysis (Pirnay-Dummer and Ifenthaler 2010; Pirnay-Dummer 2015). By this means, we compared students' essays to a reference model by using a computer-linguistic model

heuristic to track the association of concepts from a text directly to a graph. In general, different types of reference models can be used for comparison, for example, an expert's solution, a *primus-inter-pares* solution, and a “chained” reference model that is created by combining multiple expert accounts (i.e., chained expert model) or the comparably best parts of multiple learner solutions (i.e., chained *primus-inter-pares* model) (e.g., Ifenthaler 2012; Ifenthaler and Pirnay-Dummer 2011; Kopainsky et al. 2010; Lehmann et al. 2014; Lehmann and Pirnay-Dummer, 2014). In this study, the reference model was created not only combining the domain-specific expert sources presented in Table 3 to an aligned text but aggregating and integrating them into a single knowledge re-representation using the T-MITOCAR *Artemis* extension. The *Artemis* extension (Lachner and Pirnay-Dummer 2010; Pirnay-Dummer 2014; Pirnay-Dummer and Ifenthaler 2010) utilizes cluster analysis along the natural language-oriented diagnostic tools of T-MITOCAR to create an integrated model of multiple text corpora. Figure 1 illustrates the *Artemis* process and includes a miniaturized graph of the aggregated and integrated reference model.

Pirnay-Dummer and Spector (2008) suggest that analyses should incorporate those indices that correspond best with the research question and theoretical foundation. Therefore, we used the gamma matching (GAMMA) measure for the structural analysis and the balanced semantic matching (BSM) measure for the semantic-propositional analysis of students' essays. The rationale for choosing GAMMA for the structural analysis lays in the fact that this measure describes the quotient of links (edges) per concept (vertice, node) within a graph. Thus, GAMMA indices represent node density and connectedness. Spector and Koszalka (2004) acknowledge that novices' knowledge representations exhibit rather low GAMMA indices whereas expert representations show higher GAMMAs. A high matching on the GAMMA measure means that the two compared knowledge artifacts (graphs) are similarly networked—if a learner's network converges in its structure towards an expert network, this is interpreted as a sign for convergent expertise. In other words, GAMMA “indicates the cognitive structure (i.e., the breadth of understanding of the underlying subject matter)” (Lee and Spector 2012, p. 547), and it is “an appropriate measure for assessing mental models” (ibid.). Gogus (2012, 2013) also confirms that the GAMMA measure is eligible to determine a relative expertise level and to identify gaps in novices' textual re-representations. As mentioned earlier, this is particularly important in the present study in terms of assessing the degree to which learners constructed well-connected (i.e., integrated) mental models as opposed to rather fragmented models that involve separate, unconnected knowledge entities. Figure 2 presents two T-MITOCAR graphs that have been generated from essays of two students in a prior study (Lehmann and Pirnay-Dummer 2014). The graphs indicate that student A's knowledge structure (left-hand; graph A) is more integrated than student B's rather fragmented knowledge structure (right-hand; graph B). Accordingly, graph A exhibited a higher GAMMA than graph B.

Regarding the semantic analysis, we considered the BSM measure, which quantifies the semantic similarity between two knowledge representations (i.e., a student's and an expert's model), as suitable for our analysis for the following reason: BSM is the quotient of concept matching (CONC) and propositional matching (PROP) measure. It combines (a) the amount of matching concepts within models and accordingly the match in language use with (b) fully identical propositions (concept-link-concept) between a student's essay and the reference model. Hence, BSM assesses whether a student uses similar concepts and terms and links them to other concepts and terms in the same way as the reference model. Accordingly, a high BSM score is reached if both of the two compared models link concepts and terms alike when they address the same concepts/terms. Figure 3 displays two models that include matching concepts and matching propositions to illustrate



Fig. 1 Illustration of the Artemis process for the integration of the domain-specific source documents. Colors in the resulting reference model indicate knowledge domains (green: CK; red: PK; blue: PCK) (Color figure online)

the measures that result in BSM (i.e., by dividing PROP by CONC; except if CONC equals 0: then BSM is 0, too).

According to Barry et al. (2014), the BSM measure can be preferred against the other two semantic measures (i.e., CONC and PROP) “when focusing on complex causal relationships” (p. 366). Although the integration of CK, PK, and PCK is not necessarily causal,

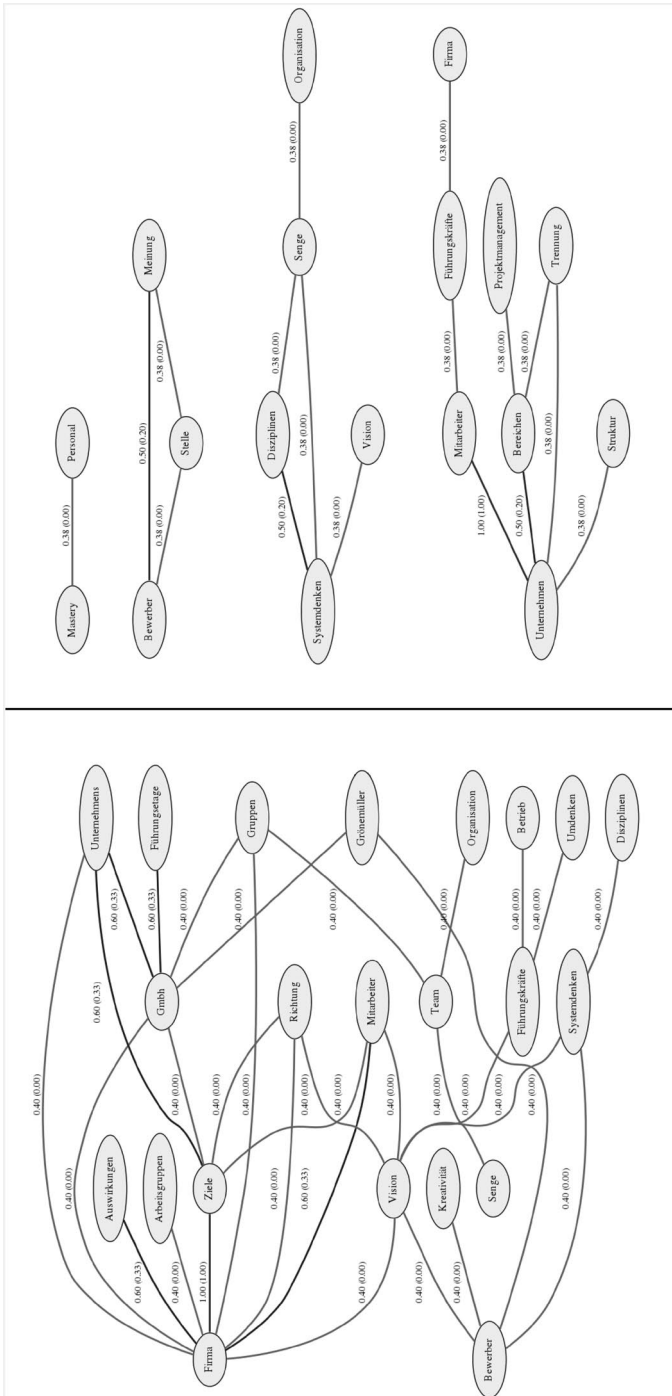


Fig. 2 T-MITOCAR graph examples with high (graph A: left-hand) and low (graph B: right-hand) GAMMA indices generated from two student essays (Lehmann and Pirnay-Dummer 2014)

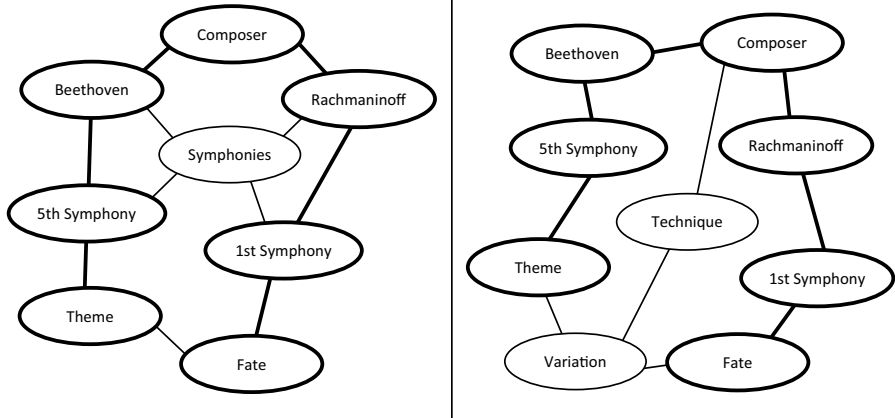


Fig. 3 Two different concept maps from music (adapted from Kopainsky et al. 2010) indicating matching concepts (heavy border weight) and matching propositions (heavy line weight)

the construction of integrated mental models through self-generated links across different domains are certainly diversified and complex, and quite a challenging task for pre-service teachers (Ball 2000; Baumert and Kunter 2013; Darling-Hammond 2006; Shulman 1987; Wäschle et al. 2015). In this study, a high BSM suggests that a student processed the learning content through “knowledge-telling” (Scardamalia and Bereiter 1991), that is, a rather passive reproduction of knowledge by referring to and linking certain concepts and terms in the exact same way as presented in the learning material. In contrast, deep knowledge integration across domains is reached through active-constructive “knowledge-transforming” (Scardamalia and Bereiter 1991; Wiley and Voss 1999), that is, by way of integrative elaborations which foster a model and level of understanding that incorporates self-generated semantic interrelations, and thus, is not just a re-sequencing of existing knowledge entities anymore (Lehmann et al. 2019).

Pirnay-Dummer et al. (2010) report good reliability coefficients (between $r = .79$ and $r = .94$), and provide further information on the two measures and their validity (which can be perceived as satisfying).

Results

Descriptive statistics for the structural and semantic T-MITOCAR measures used in the present study (i.e., GAMMA and BSM) are reported in Table 4.

Table 4 Means and standard deviations for the GAMMA and BSM measure for each experimental group

Level of measurement	Measure	Experimental group			
		UWT	UWT + P	AWT	AWT + P
		<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>
Structural	GAMMA	.54 (.28)	.68 (.15)	.61 (.17)	.63 (.14)
Semantic	BSM	.21 (.20)	.15 (.16)	.18 (.18)	.14 (.15)

We conducted a multivariate omnibus test (i.e., MANOVA) with the two dependent measures GAMMA and BSM, and the two between-subjects factors ‘writing task’ and ‘prompts’ to check our hypotheses. The omnibus MANOVA indicated that participants’ knowledge integration as assessed by GAMMA and BSM was not affected by the writing task, Wilks’ $\lambda = .995$, $F(2,87) = .208$, $p > .05$, $\eta_p^2 = .005$. However, there was a significant main effect of prompts on the combined GAMMA and BSM measures, Wilks’ $\lambda = .895$, $F(2,87) = 5.124$, $p = .008$, $\eta_p^2 = .105$. No significant interaction effect was found, Wilks’ $\lambda = .969$, $F(2,87) = 1.396$, $p > .05$, $\eta_p^2 = .031$.

Following the significant main effect of prompts identified by the multivariate omnibus test, two ANOVAs were applied to estimate the effect of prompts on (a) the structural measure (i.e., GAMMA) and (b) the semantic measure (i.e., BSM) separately. Results showed a significant effect of prompts on the GAMMA measure, $F(1,88) = 4.439$, $p = .038$, $\eta_p^2 = .048$. Essays of students from experimental conditions with prompts ($M = .658$, $SE = .028$) were more comparable to the reference model in terms of the density of vertices than essays of students from unprompted conditions ($M = .574$, $SE = .028$). No significant differences were found on the semantic level as measured by BSM, $F(1,88) = 1.855$, $p = .177$, $\eta_p^2 = .021$.

Discussion

To become an effective teacher, pre-service teachers need to not only acquire knowledge of different domains, that is, content knowledge (CK) of the subject that will be taught later on, subject-matter specific pedagogical content knowledge (PCK), and generic pedagogical knowledge that transfers to teaching irrespective of the subject to be taught (Baumert and Kunter 2013; Darling-Hammond 2006; Shulman 1986, 1987); they also have to integrate their pre-existing and newly acquired knowledge within and across topics and domains into a more coherent interrelated model to develop expertise (Berliner 2001; Bromme 2014; Livingston and Borko 1990). This, however, is not an easy task to accomplish, especially since pre-service teacher education programs usually are fragmented and not systematically linked across the core domains of teachers’ professional knowledge (i.e., CK, PK, and PCK; Darling-Hammond 2006). Yet, recent findings on the integration of information and knowledge entities from various sources suggest that multiple document learning environments foster the construction of integrated deep comprehension knowledge representations (e.g., Bråten and Strømsø 2009; Gil et al. 2010; Wiley and Voss 1996, 1999). However, it remained open whether this could also apply to pre-service teachers’ knowledge integration since integrating knowledge from conceptually distinct domains might be complex task to accomplish. As a result, some researchers (e.g., Lehmann et al. 2019; Wäschle et al. 2015) took first steps to investigate instructionally guided multiple-document learning environments that were designed to stimulate pre-service teachers’ integration of CK, PK, and PCK. First empirical evidence for the efficacy of writing-to-learn tasks and prompts in enhancing knowledge integration *across* domains was found.

In accordance with this line of research, the present study aimed at further investigating such instructional approaches (i.e., tasks and prompts) for pre-service teacher education. Moreover, the study sought to explore the potential of the automated assessment and analysis tool T-MITOCAR in predicting separative and integrative information processing in verbally externalized mental models. To this end, we used specific T-MITOCAR measures to find out whether our writing tasks for global comprehension (unspecific vs.

argumentative) and the provision of integration prompts (five task-supplemental focus questions) differently affected learners' construction of an integrated mental models of different knowledge domains, or rather the re-representation of the model. In terms of knowledge structure we used the quotient of links per concept within externalized models in comparison to a reference model (i.e., GAMMA matching). With regard to semantic we used a measure which combines the amount of matching concepts within models and accordingly the match in language use with fully identical propositions as compared to the reference model (i.e., BSM matching).

Regarding our first hypothesis, which was focused on the structural GAMMA measure, results of the study supported that task-supplemental focus questions are effective prompts for structural knowledge integration (Hypothesis H1a). The texts written by the students from the prompted condition were more likely to inherit a similar networked structure to the expert model than the texts of the students who did not receive prompts. Thus, the prompts improved the construction of integrated mental models in terms of knowledge integration *across* different domains (i.e., CK, PK, PCK). This finding is consistent with prior work on instructional approaches that are designed to "prompt" learning processes that are conducive for knowledge integration when learning from multiple texts (e.g., McCrudden and Schraw 2007, 2010; Wäschle et al. 2015). In terms of theory, this finding can be interpreted against the creation of an adequate task model (Rouet and Britt 2011), which gears a learner's information processing strategies towards structural cross-relations. Accordingly, task-supplemental cues induce a better understanding of the task at hand, which subsequently shifts learners' default mental modeling to a more strategic process of searching, evaluating, and structurally integrating domain-specific information and knowledge entities derived from various learning material.

Our second hypothesis on the effects of prompts addressed knowledge semantics in terms of fully identical proposition matches (BSM measure). We assumed that externalized models of students who did not receive integration prompts in the form of guiding questions depict a higher semantic-propositional similarity with the reference model because they process information less transformative (Hypothesis H2a). However, there were no significant differences between the experimental conditions. Hence, the integration prompts did not affect knowledge semantics in constructing integrated mental models. More research will be needed to gain a comprehensive insight into pre-service teachers' integration across CK, PK, and PCK on the semantic level.

Moreover, this study has been unable to provide further evidence that argument writing tasks facilitate more integrated mental models of multiple domain-specific texts (Hypotheses H1b and H2b). This is somewhat surprising since argument tasks have been found to be particularly effective in enhancing knowledge integration when learning with multiple documents (e.g., Bråten and Strømsø 2009; Lehmann et al. 2019; Wiley and Voss 1996, 1999). Thus, several questions arise. First, it remains open under which circumstances argument tasks lead to better performances in the construction of integrated mental models across conceptually distinct domains in multiple-document learning environments. Gil et al. (2010) suggests to take individual characteristics of the learner into account, for example, epistemological beliefs, pre-existing topic knowledge, and argument writing skills. Second, these results put into question in what way the writing tasks influenced not the construction but the externalization of mental models, and whether writing an argument limited students in re-representing their inferred integrative knowledge transformation on a semantic level. More generally spoken, different forms of representation may yield different results. Hence, we suggest that additional studies consider students' language skills when using natural language based assessment, and their experience with other forms of representation

that are used for model-based analysis (e.g., concept mapping). Finally, since written composition (a) re-represents knowledge structure and semantics, rhetoric, and pragmatics (Halford et al. 2010; Leech 1983; van Dijk 1977), and (b) involves two interacting problem spaces (i.e., the content problem space, which includes structure and semantics, and the rhetorical problem space, which includes thoughts about how to adequately represent an internal model) it remains unanswered to what degree participants' rhetoric and pragmatics restricted the comparison of their essays with the reference model (see also limitations below).

Of course, our study also comes with certain limitations. First, mental models are not stored permanently in long term memory. They are characterized to be created ad hoc, thus, they have only limited stability (Seel 1991). Therefore, the externalized models that served as a basis for the analysis of participants' knowledge integration across multiple domains do not allow to draw inferences about well-established knowledge structures such as schemas. But pre-service teachers need to develop schemas, because they regulate actions in a top-down manner (instead of bottom-up), thus, allowing immediate decision-making, for example, in complex classroom situations. Having said that, it is also important to note that the recurring construction of similar mental models is assumed to foster rather stable and persistent cognitive structures. Hence, our study makes an important contribution to pre-service teachers' knowledge integration as a form of learning, even though it is limited with regard to the question of a learning-dependent transition of mental models to a schema. Our second limitation is that the reference model was created by the T-MITOCAR Artemis extension, which aggregated and integrated the domain-specific text documents (which served as learning material for the participants). Although this truly represented an integrated solution, it was not an argumentative text. Depending on the configuration of the reference model we might be able to reach different results. Hence, we suggest that a future study incorporates different types of expert solutions as reference models for the analysis. For example, teaching experts, both academic and practicing, could perform the writing tasks and their solutions could then be used as reference models. Also, these set of solutions could be framed into an integrated chained reference model. Especially with regard to argumentative essays, this could provide further insights into the construction of integrated mental models of multiple domain-specific instructional texts due to their different text genre, which is likely to influence knowledge semantics, rhetoric, and pragmatics. Another possibility is to compose the reference model for analysis collaboratively by multiple experts, one of each involved domain (e.g., CK, PK, and PCK). Finally, more CK and PCK domains are needed to be able to generalize the present findings on the integration of teachers' professional knowledge domains since knowledge integration across these domains might be affected by the learning content to be integrated, too (Hashweh 2005). Also, the consideration of different topics within the same knowledge domains could be fruitful to deepen our understanding of knowledge integration in dependence of various topic characteristics.

Conclusion

In terms of multiple-documents learning environments for initial pre-service teacher education, this study showed that integration prompts (i.e., task-supplemental focus questions) stimulate students' structural integration of knowledge from different domains into a common mental model (H1). This finding is in line with other studies that found specific

prompts to be an effective instructional tool to shift students' cognitive processing from a rather passive (separative) knowledge-telling of domain-specific information to a more integrative process of knowledge-transforming (e.g., Lehmann et al. 2019; Scardamalia and Bereiter 1991; Wäschle et al. 2015). Yet, prompts did not affect knowledge semantics in constructing integrated mental models across multiple domains (H2). With regard to argument writing tasks, there were no effects found by way of the model-based analyses. This is surprising considering that argument writing tasks have been repeatedly found to stimulate students' knowledge integration effectively (e.g., Bråten and Strømsø 2009; Wiley and Voss 1999). However, different measures have been used in those studies. As discussed above, a possible explanation may rest in the reference model, which was not set up as an argumentative essay. More studies are needed to further explore whether a differently composed reference model will lead to different findings. Also, a comparison of different knowledge integration indicators, for example, measures from category-driven content analyses and from automated model-based analyses, might help to establish a greater degree of accuracy on assessing knowledge integration across domains.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Acknowledgements The model-based analyses and results presented in this article are based on data collected but not reported by Lehmann et al. (2019).

Funding This research was supported in the course of the University of Bremen's future concept by the Excellence Initiative of the German federal and state governments.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Ball, D. L. (2000). Bridging practices. Intertwining content and pedagogy in teaching and learning to teach. *Journal of Teacher Education*, 51(3), 241–247. <https://doi.org/10.1177/0022487100051003013>.
- Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching. What makes it special? *Journal of Teacher Education*, 59(5), 389–407. <https://doi.org/10.1177/0022487108324554>.
- 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>.
- Barry, D., Bender, N., Breuer, K., & Ifenthaler, D. (2014). Shared cognitions in a field of informal learning. Knowledge maps towards money management of young adults. In D. Ifenthaler & R. Hanewald (Eds.), *Digital knowledge maps in education. Technology enhanced support for teachers and learners* (pp. 355–370). New York: Springer.
- Baumert, J., & Kunter, M. (2013). Professionelle Kompetenz von Lehrkräften [Professional competence of teachers]. In I. Gogolin, H. Kuper, H. H. Krüger, & J. Baumert (Eds.), *Stichwort: Zeitschrift für Erziehungswissenschaft* (pp. 277–337). Wiesbaden: Springer. https://doi.org/10.1007/978-3-658-00908-3_13.
- 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](https://doi.org/10.1016/S0883-0355(02)00004-6).

- Bråten, I., & Strømsø, H. I. (2009). Effects of task instruction and personal epistemology on the understanding of multiple texts about climate change. *Discourse Processes*, 47(1), 1–37. <https://doi.org/10.1080/01638530902959646>.
- Bromme, R. (2014). *Der Lehrer als Experte. Zur Psychologie des professionellen Wissens [The teacher as an expert. On the psychology of professional knowledge]*. Münster: Waxmann Verlag.
- Bruner, J. S. (1964). The course of cognitive growth. *American Psychologist*, 19(1), 1–15.
- Brunner, E. (2014). *Mathematisches Argumentieren, Begründen und Beweisen. Grundlagen, Befunde und Konzepte [Mathematical reasoning, justifying, and proving. Basics, findings, and concepts]*. Berlin: Springer Spektrum.
- Chi, M. T. H., de Leeuw, N., Chiu, M.-H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, 18(3), 439–477. https://doi.org/10.1207/s15516709cog1803_3.
- Clariana, R. B. (2010). Deriving individual and group knowledge structure from network diagrams and from essays. In D. Ifenthaler, P. Pirnay-Dummer, & N. M. Seel (Eds.), *Computer-based diagnostics and systematic analysis of knowledge* (pp. 117–130). New York: Springer.
- 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. (1990). *Representations of commonsense knowledge*. San Mateo: Morgan Kaufmann.
- Frazier, L. (1999). *On sentence interpretation*. Dordrecht: Kluwer.
- 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>.
- Gogus, A. (2012). Evaluation of mental models: Using Highly Interactive Model-based Assessment Tools and Technologies (HIMATT) in mathematics domain. *Technology, Instruction, Cognition and Learning*, 9(1), 31–50.
- Gogus, A. (2013). Evaluating mental models in mathematics: A comparison of methods. *Educational Technology Research and Development*, 61(2), 171–195. <https://doi.org/10.1007/s11423-012-9281-2>.
- Grieser, D. (2015). *Analysis I*. Wiesbaden: Springer.
- Halford, G. S., Wilson, W. H., & Phillips, S. (2010). Relational knowledge: The foundation of higher cognition. *Trends in Cognitive Sciences*, 14(11), 497–505. <https://doi.org/10.1016/j.tics.2010.08.005>.
- 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>.
- Harr, N., Eichler, A., & Renkl, A. (2015). Integrated learning: Ways of fostering the applicability of teachers' pedagogical and psychological knowledge. *Frontiers in Psychology*, 6, 1–16. <https://doi.org/10.3389/fpsyg.2015.00738>.
- Hashweh, M. Z. (1987). Effects of subject-matter knowledge in the teaching of biology and physics. *Teaching & Teacher Education*, 3(2), 109–120. [https://doi.org/10.1016/0742-051X\(87\)90012-6](https://doi.org/10.1016/0742-051X(87)90012-6).
- Hashweh, M. Z. (2005). Teacher pedagogical constructions: A reconfiguration of pedagogical content knowledge. *Teachers and Teaching: Theory and Practice*, 11(3), 273–292. <https://doi.org/10.1080/13450600500105502>.
- Ifenthaler, D. (2010). Relational, structural, and semantic analysis of graphical representations and concept maps. *Educational Technology Research and Development*, 58(1), 81–97. <https://doi.org/10.1007/s11423-008-9087-4>.
- Ifenthaler, D. (2011). Identifying cross-domain distinguishing features of cognitive structures. *Educational Technology Research and Development*, 59(6), 817–840. <https://doi.org/10.1007/s11423-011-9207-4>.
- Ifenthaler, D. (2012). Determining the effectiveness of prompts for self-regulated learning in problem-solving scenarios. *Journal of Educational Technology & Society*, 15(1), 38–52.
- Ifenthaler, D., & Pirnay-Dummer, P. (2011). States and processes of learning communities. Engaging students in meaningful reflection and learning. In B. White, I. King, & P. Tsang (Eds.), *Social media tools and platforms in learning environments* (pp. 81–94). Berlin: Springer. <https://doi.org/10.1007/978-3-642-20392-3>.
- Ifenthaler, D., & Pirnay-Dummer, P. (2014). Model-based tools for knowledge assessment. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of research on educational communications and technology* (4th ed., pp. 289–301). New York: Springer. <https://doi.org/10.1007/978-1-4614-3185-5>.
- 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>.
- Johnson, T. E., Ifenthaler, D., Pirnay-Dummer, P., & Spector, J. M. (2009). Using concept maps to assess individuals and team in collaborative learning environments. In P. L. Torres & R. C. V. Marriott

- (Eds.), *Handbook of research on collaborative learning using concept mapping* (pp. 358–381). Hershey, PA: Information Science Publishing.
- Johnson-Laird, P. N. (1983). *Mental models: Toward a cognitive science of language, inference and consciousness*. Boston, MA: Harvard University Press.
- Kintsch, W. (1974). *The representation of meaning in memory*. New York: Erlbaum.
- Kopainsky, B., Pirnay-Dummer, P., & Alessi, S. M. (2010). Automated assessment of learners' understanding in complex dynamic systems. In *Proceedings of the 28th international conference of the system dynamics society*, Seoul.
- Kopp, K. J. (2013). *Selecting and using information from multiple documents for argumentation*. Doctoral dissertation. Northern Illinois University. Ann Arbor, MI: ProQuest LLC.
- Krauss, S., Brunner, M., Kunter, M., Baumert, J., Blum, W., Neubrand, J., et al. (2008). Pedagogical content knowledge and content knowledge of secondary mathematics teachers. *Journal of Educational Psychology, 100*(3), 716–725. <https://doi.org/10.1037/0022-0663.100.3.716>.
- Lachner, A., & Pirnay-Dummer, P. (2010). Model-based knowledge mapping—A new approach for the automated graphical representation of organizational knowledge. In J. M. Spector, D. Ifenthaler, P. Isaias, Kinshuk, & D. G. Sampson (Eds.), *Learning and instruction in the digital age: Making a difference through cognitive approaches, technology-facilitated collaboration and assessment, and personalized communications* (pp. 69–85). New York: Springer.
- Le Ny, J.-F. (1993). Wie kann man mentale Repräsentationen repräsentieren? [How to represent mental representations?]. In J. Engelkamp & T. Pechmann (Eds.), *Mentale Repräsentation [Mental representation]* (pp. 31–39). Bern: Huber.
- Lee, J., & Spector, J. M. (2012). Effects of model-centered instruction on effectiveness, efficiency, and engagement with ill-structured problem solving. *Instructional Science, 40*(3), 537–557. <https://doi.org/10.1007/s11251-011-9189-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>.
- Leech, G. N. (1983). *Principles of pragmatics*. New York: Routledge.
- Lehmann, T., Hähnlein, I., & Ifenthaler, D. (2014). Cognitive, metacognitive and motivational perspectives on reflection in self-regulated online learning. *Computers in Human Behavior, 32*, 313–323. <https://doi.org/10.1016/j.chb.2013.07.051>.
- Lehmann, T., & Pirnay-Dummer, P. (2014). *Expertise divergence and convergence development in complex problem-based learning scenarios*. Poster presentation at the annual meeting of the American Educational Research Association (AERA), Philadelphia, PA, USA.
- 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>.
- Lewin, K. (1922). Das Problem der Wissensmessung und das Grundgesetz der Assoziation. Teil 1 [The problem of knowledge assessment and the basic law of association]. *Psychologische Forschung, 1*, 191–302. <https://doi.org/10.1007/BF00410391>.
- Linn, M. C. (2000). Designing the knowledge integration environment. *International Journal of Science Education, 22*, 781–796. <https://doi.org/10.1080/095006900412275>.
- List, A., & Alexander, P. A. (2017). Analyzing and integrating models of multiple text comprehension. *Educational Psychologist, 52*(3), 143–147. <https://doi.org/10.1080/00461520.2017.1328309>.
- 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>.
- 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>.
- McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review, 99*(3), 440–466. <https://doi.org/10.1037/0033-295X.99.3.440>.
- McNamara, T. P. (1992). Priming and constraints it places on theories of memory and retrieval. *Psychological Review, 99*(4), 650–662. <https://doi.org/10.1037/0033-295X.99.4.650>.
- McNamara, T. P. (1994). Priming and theories of memory: A reply to Ratcliff and McKoon. *Psychological Review, 101*(1), 185–187. <https://doi.org/10.1037/0033-295X.101.1.185>.
- McNamara, T. P., Miller, D. L., & Bransford, J. D. (1991). Mental models and reading comprehension. In R. Barr, M. Kamil, P. Mosenthal, & P. Pearson (Eds.), *Handbook of reading research* (Vol. 2, pp. 490–511). London: Longman.

- Perfetti, C. A., Rouet, J.-F., & Britt, M. A. (1999). Toward a theory of documents representation. In H. van Oostendorp & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 99–122). Mahwah, NJ: Lawrence Erlbaum Associates.
- Pirnay-Dummer, P. (2014). Gainfully guided misconception. How automatically generated knowledge maps can help companies within and across their projects. In D. Ifenthaler & R. Hanewald (Eds.), *Digital knowledge maps in higher education. Technology-enhanced support for teachers and learners* (pp. 253–274). New York: Springer. https://doi.org/10.1007/978-1-4614-3178-7_14.
- Pirnay-Dummer, P. (2015). Linguistic analysis tools. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 427–442). New York: Guilford Publications.
- Pirnay-Dummer, P., & Ifenthaler, D. (2010). Automated knowledge visualization and assessment. In D. Ifenthaler, P. Pirnay-Dummer, & N. M. Seel (Eds.), *Computer-based diagnostics and systematic analysis of knowledge* (pp. 77–115). New York: Springer. https://doi.org/10.1007/978-1-4419-5662-0_6.
- Pirnay-Dummer, P., Ifenthaler, D., & Spector, J. M. (2010). Highly integrated model assessment technology and tools. *Educational Technology Research and Development*, 58(1), 3–18. <https://doi.org/10.1007/s11423-009-9119-8>.
- Pirnay-Dummer, P., & Spector, J. M. (2008). *Language, association, and model re-representation. How features of language and human association can be utilized for automated knowledge assessment*. Paper presented at the annual meeting of the AERA, New York, NY, USA.
- Pollio, H. R. (1966). *The structural basis of word association behavior*. The Hague: Mouton.
- Ratcliff, R., & McKoon, G. (1994). Retrieving information from memory: Spreading-activation theories versus compound-cue theories. *Psychological Review*, 101(1), 177–184. <https://doi.org/10.1037/0033-295X.101.1.177>.
- 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. (2015). Wissenserwerb [Knowledge acquisition]. In E. Wild & J. Möller (Eds.), *Pädagogische Psychologie [Educational psychology]* (pp. 3–24). Berlin: Springer.
- Renkl, A., Mandl, H., & Gruber, H. (1996). Inert knowledge: Analyses and remedies. *Educational Psychologist*, 31(2), 115–121. https://doi.org/10.1207/s15326985Sep3102_3.
- 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: Information Age Publishing.
- Rouet, J.-F., Britt, M. A., & Durik, A. M. (2017). RESOLV: Readers' representation of reading contexts and tasks. *Educational Psychologist*, 52(3), 200–215. <https://doi.org/10.1080/00461520.2017.1329015>.
- Russel, W. A., & Jenkins, J. J. (1954). *The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff word association test*. Technological Report 11, University of Minnesota.
- Scardamalia, M., & Bereiter, C. (1987). *The psychology of written composition*. New York: Routledge.
- Scardamalia, M., & Bereiter, C. (1991). Literate expertise. In K. A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise. Prospects and limits* (pp. 172–194). Cambridge: University Press.
- Seel, N. M. (1991). *Weltwissen und mentale Modelle [World knowledge and mental models]*. Göttingen: Hogrefe.
- Seel, A. (1997). Von der Unterrichtsplanung zum konkreten Lehrerhandeln [From lesson planning to practical teaching]. *Unterrichtswissenschaft*, 25(3), 257–273.
- Seel, N. M. (1999). Educational semiotics: School learning reconsidered. *Journal of Structural Learning & Intelligent Systems*, 14(1), 11–28.
- Seufert, T., Zander, S., & Brünken, R. (2007). Das Generieren von Bildern als Verstehenshilfe beim Lernen aus Texten [Generating pictures as aid for comprehension in learning from texts]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 39(1), 33–42. <https://doi.org/10.1026/0049-8637.39.1.33>.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14. <https://doi.org/10.3102/0013189X015002004>.
- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57(1), 1–22. <https://doi.org/10.3102/0013189X015002004>.
- Shute, V. J., Masduki, I., Donmez, O., Kim, Y. J., Dennen, V. P., Jeong, A. C., et al. (2010). Assessing key competencies within game environments. In D. Ifenthaler, P. Pirnay-Dummer, & N. M. Seel (Eds.), *Computer-based diagnostics and systematic analysis of knowledge* (pp. 281–309). New York: Springer.
- Spector, J. M., Dennen, V. P., & Koszalka, T. A. (2005). Individual and collaborative construction of causal concept maps: An online technique for learning and assessment. In G. Chiazzese, M. Allegra, A. Chifari, & S. Ottaviano (Eds.), *Methods and technologies for learning* (pp. 223–227). Southampton: WIT-Press.
- Spector, J. M., & Koszalka, T. A. (2004). *The DEEP methodology for assessing learning in complex domains (Final report the National Science Foundation Evaluative Research and Evaluation Capacity Building)*. Syracuse, NY: Syracuse University.

- Stachowiak, F. J. (1979). *Zur semantischen Struktur des subjektiven Lexikons [On the semantic structure of the subjective lexicon]*. München: Wilhelm Fink Verlag.
- Tynjälä, P., Mason, L., & Lonka, K. (Eds.). (2001). *Writing as a learning tool: Integrating theory and practice*. Dordrecht: Kluwer Academic Publishers. <https://doi.org/10.1007/978-94-010-0740-5>.
- van Dijk, T. A. (1977). *Text and context: Explorations in the semantics and pragmatics of discourse*. New York: Longman.
- van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.
- Verloop, N., Van Driel, J., & Meijer, P. (2001). Teacher knowledge and the knowledge base of teaching. *International Journal of Educational Research*, 35(5), 441–461. [https://doi.org/10.1016/S0883-0355\(02\)00003-4](https://doi.org/10.1016/S0883-0355(02)00003-4).
- Voss, T., Kunter, M., & Baumert, J. (2011). Assessing teacher candidates' general pedagogical/psychological knowledge: Test construction and validation. *Journal of Educational Psychology*, 103(4), 952–969. <https://doi.org/10.1037/a0025125>.
- 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. (1996). The effects of 'playing historian' on learning in history. *Applied Cognitive Psychology*, 10(7), 63–72. [https://doi.org/10.1002/\(SICI\)1099-0720\(199611\)10:7<63::AID-ACP438>3.0.CO;2-5](https://doi.org/10.1002/(SICI)1099-0720(199611)10:7<63::AID-ACP438>3.0.CO;2-5).
- 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>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Thomas Lehmann is a postdoctoral researcher at the University of Bremen's research unit "Learning, Instruction, and Educational Psychology". After graduating in 2011 (M.A. in Educational Science), he completed his PhD (Dr. phil.) in 2014 at the University of Freiburg, Germany. His dissertation dealt with the interaction of cognition and emotion in problem solving. His current research focuses on knowledge integration, mental models, self-regulated learning, decision making, problem solving, and instructional design. He teaches courses on research methods, pedagogical-psychological diagnostics, and the design of learning environments. Besides, he represents the Faculty for Pedagogy and Educational Sciences in the University's advisory committee for research, young academics, and transfer.

Pablo Pirnay-Dummer is full professor of Educational Psychology at the Martin-Luther-University of Halle-Wittenberg. He graduated from the University of Freiburg, Germany. His research and publications are located in the area of language, cognition, learning, expertise and technology with a particular interest on the use of language in learning processes. Also, some of his applied work is on model based knowledge management and organizational learning. He developed computer-linguistic language-oriented knowledge assessment methodologies that are built to assess, analyse and compare individual and group models of expertise. Among other research interests, he also developed the web based training framework L-MoSim (Learner Model Simulation), including new approaches of automated tasks synthesis for complex problem solving.

Florian Schmidt-Borcherding is a full professor for Learning, Instruction, and Educational Psychology at the Faculty for Pedagogical and Educational Sciences, University of Bremen. His research interests include instructional design, multimedia learning, educational videos, self-regulated learning, and measures of learning processes. He also teaches courses on research methods, pedagogical-psychological diagnostics, and learning and instruction for both pre-service teachers and prospective educational specialists.