Guiding Knowledge Exchange in Collaborative Learning: Mechanisms and Potential of Text-Mining Support

Von der Fakultät für Ingenieurwissenschaften,
Abteilung Informatik und Angewandte Kognitionswissenschaft
der Universität Duisburg-Essen

zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften (Dr. rer. nat.)

genehmigte kumulative Dissertation

von

Melanie Erkens
aus
Linnich, Deutschland

Erstgutachter: Prof. Dr. Daniel Bodemer
Zweitgutachter: Prof. Dr. Armin Weinberger
Tag der mündlichen Prüfung: 21.05.2019

Acknowledgements

After four years, today is the day: writing this note of thanks is the finishing touch on my dissertation. Although I am very happy to bring this project to completion, I will always appreciate this significant phase of my life, which has been a time of intense learning both in the scientific area and in the personal realm. I will be eternally grateful for this opportunity that has been given to me, and I would like to reflect on the people who have supported and considerably helped me throughout this period.

First, I would like to thank my colleagues from the Institute for Research Methods in Psychology–Media-based Knowledge Construction for their wonderful collaboration. Knowledge is considered the only resource that grows through sharing—I thank you for letting it continuously grow. You strongly supported me and constantly provided assistance and motivation. I would like to single out my supervisor at the lab, Prof. Dr. Daniel "the sense maker" Bodemer. Thank you Daniel for your excellent guidance and for all of the opportunities to conduct my research and dissertation at the University of Duisburg-Essen. I have learned immensely from you—and I still do. Becoming a part of your team is indeed a blessing; you not only enable research at a high level but also foster a fantastic team spirit. I would also like to express my gratitude to Lenka Schnaubert and Sven Heimbuch. Lenka, you are gold—wise, passionate about science, and readily available when help is necessary. Sven, thank you for the liters of freshly brewed coffee, your statistical support, and all of the nerdish matters we engaged in during our breaks.

My sincere gratitude also goes to Prof. Dr. Heinz Ulrich Hoppe who provided valuable guidance at the beginning of my journey. Next to Daniel, you were the one who gave impetus to my project. You equipped me with the tools for choosing the right direction.

Finally, I would like to convey my deepest appreciation to family and friends, foremost of whom are my aunt Hildegard Gillessen and my uncle Konrad Gillessen. Without the two of you, I would not have been able to pursue this path—thank you for your interest in my work, your unwavering attention, and your priceless support. I am likewise profoundly thankful to Benjamin Walter. Without you, the little Pilla and the little Piff, this time would not have been so good to master. Nobody could have given me more enjoyably escapes from work. Last but not the least, I wish to thank my friends. I truly appreciate our great times together and the energy you give me. Girls, you are indeed the best!

Once again, thank you very much, everyone! Melanie Erkens, Cologne, January 31, 2019

Abstract

English

Knowledge exchange is a powerful asset for promoting learning, but not every exchange of knowledge improves learning outcomes; to this purpose, collaborative learning offers instructional guidance. In particular, instructors can form groups of learners with complementary cognitive characteristics (group formation) or support their awareness of the learning partners' characteristics (group awareness support), as both measures engage people to think or communicate in learning-enhancing ways. Therefore, instructors need to collect information about the learners as an initial step, which can involve some effort. Text-mining methods seem suitable for reducing this effort because they can automatically extract the information required from available learner-generated text. Moreover, this bottom-up approach introduces new means of activating prior knowledge by providing the collected information to the learners. To realize this promising potential, however, instructional guidance and sophisticated methods from computer science need to be integrated and systematically explored under consideration of the underlying mechanisms.

This work approaches this goal through three studies. Study 1 compares different text-mining methods in terms of their suitability for identifying text differences (for grouping) and extracting cognitive information (for group awareness support) from a text corpus with predefined content properties. The resulting selection informs the design of the integrated *Grouping and Representing Tool*. Applying this tool, Study 2 investigates the effect of text mining-based guidance on learning outcomes in a classroom setting. Results indicate that students learn better and converge their knowledge more when this tool supports them. In addition to verifying this general effect of text mining-based guidance, Study 3 disentangles the guiding effects of the provided information about learning partners and learning content in a systematic laboratory setting with simulated text mining. Although the results do not confirm the improvement of prior knowledge activation, they reveal the new finding that information about learning partners should be provided without content-specification to support cognitive elaboration and that its effect on learning is mediated by knowledge integration and partner modeling accuracy. Furthermore, the results suggest that the effects of the provided information on cognitive, metacognitive, and communication processes

found in previous research can also be induced by providing text mining-generated information.

The three studies included in this dissertation contribute to the research area of computer-supported collaborative learning by exploring the usefulness of text-mining methods in knowledge exchange. In particular, the designed tool for text mining-based grouping and representing can be used as a group awareness tool for enhancing the instructors' efficiency. Single functions thereof can also be used in other scenarios where learners produce texts (e.g., individual learning settings) to relieve instructors. Moreover, new insights into mechanisms triggered by the provided information types, whether text mining-gathered or not, can enrich other areas of (technology-enhanced) learning. They can improve tool designs or can be applied to other learning scenarios where cognitive, metacognitive, or communication processes are to be supported.

Deutsch

Der Austausch von Wissen ist fundamental fürs Lernen. Allerdings verbessert nicht jeder Wissensaustausch die Lernergebnisse; oftmals benötigen Lernende zusätzliche Unterstützung, die durch Instruktionen aus dem Bereich des kollaborativen Lernens gewährleistet werden kann. Dabei haben sich zwei Ansätze besonders bewährt: Lehrer können Lerngruppen bilden (Group formation), in denen die kognitiven Eigenschaften der Gruppenmitglieder auf bestimmte Weise verteilt sind, oder sie verschaffen den Lernern Kenntnisse über die kognitiven Eigenschaften ihrer Lernpartner (Group awareness support). Beides geht mit verbesserten Denk- wie Kommunikationsprozessen einher. Voraussetzung zur Umsetzung dieser Maßnahmen ist jedoch, dass die Lehrer Informationen über die Lernenden besitzen, deren Einholung mit hohem Aufwand verbunden sein kann. Um diesen Aufwand zu reduzieren, bieten sich Text-Mining-Methoden an, die die benötigten Informationen automatisch aus vorhandenen Lerner-Texten extrahieren können. Zudem verspricht der Bottom-up-Ansatz dieser Erfassung, Vorwissen besser aktivieren zu können, wenn die so generierten Informationen an die Lernenden zurückgemeldet werden. Um dieses Potenzial auszuschöpfen, sind jedoch zunächst Instruktionsdesign und Informatik-Funktionen zu integrieren und systematisch unter Berücksichtigung zugrundeliegender Wirkmechanismen zu erforschen.

Diesem Ziel nähert sich die vorliegende Arbeit mit drei Studien. Studie 1 vergleicht verschiedene Text-Mining-Methoden hinsichtlich ihrer Eignung, Textunterschiede (für die Gruppierung) und kognitive Informationen (zur Unterstützung der besseren Kenntnis der Lernpartner) aus einem Textkorpus mit vordefinierten Inhalten zu extrahieren. Die hierauf basierend ausgewählten Methoden dienen dem Design eines integrierten *Grouping and Representing Tools*. Unter Einsatz dieses Tools untersucht Studie 2 dessen Effekt auf Lernergebnisse durch den Wissensaustausch im Schulunterricht. Die Ergebnisse zeigen, dass Schüler besser lernen und ihr Wissen stärker aneinander annähern, wenn sie von dem Tool unterstützt werden. Neben dieser Bestätigung eines generellen Effekts Text-Mining-basierter Anleitung fokussiert Studie 3 die Untersuchung von Einzel- und Interaktionseffekten der bereitgestellten Informationen über Lernpartner und Lerninhalte in einer systematischen Laborstudie mit simuliertem Text Mining. Obwohl sich die Verbesserung der Vorwissensaktivierung nicht bestätigt, liefert die Studie die neue Erkenntnis, dass Informationen über Lernpartner zur Verbesserung von kognitiver Elaboration inhaltlich nicht zu feingliedrig bereitgestellt werden sollten, und dass die Wirkung solcher Informationen auf das Lernergebnis durch verbesserte Wissensintegration und genauere Partnereinschätzung vermittelt wird. Darüber hinaus deuten die Ergebnisse darauf hin, dass lernförderliche Effekte der bereitgestellten Informationen auf kognitive, metakognitive wie auch Kommunikationsprozesse, die in früherer Forschung identifiziert wurden, auch durch die Bereitstellung von Informationen auf Basis der Text-Mining-Simulation verursacht werden können.

Die drei in dieser Dissertation enthaltenen Studien tragen zur Forschung im Bereich des kollaborativen Lernens bei, indem sie den Nutzen von Text-Mining-Methoden für den Wissensaustausch untersuchen und abstecken. Das im Zuge dessen entwickelte Tool kann für die Text-Mining-basierte Gruppierung Lernender und die Rückmeldung von Informationen zur Verbesserung der Kenntnis des Lernpartners verwendet werden, um so die Arbeit der Lehrer zu erleichtern. Einzelne Funktionen des Tools sind zur Lehrerentlastung auch in anderen Szenarien denkbar, in denen Lernende Texte produzieren, z.B. auch in individuellen Lernszenarien. Darüber hinaus können die neuen Erkenntnisse über Mechanismen, die durch die bereitgestellten Informationstypen ausgelöst werden, unabhängig davon, ob sie per Text Mining erfasst wurden oder nicht, andere Bereiche des (technologiebasierten) Lernens bereichern. Sie können zur Verbesserung von Tools dienen oder auf andere Lernszenarien angewendet werden, in denen kognitive, metakognitive oder Kommunikationsprozesse unterstützt werden sollen.

Index

List of included papers

Paper 1

Erkens, M., Bodemer, D., & Hoppe, H. U. (2016). Improving collaborative learning in the classroom: Text mining based grouping and representing. *International Journal of Computer-Supported Collaborative Learning, 11*(4), 387–415. doi:10.1007/s11412-016-9243-5

→ Studies 1 and 2

Paper 2

Erkens, M., & Bodemer, D. (2017). Which visualization guides learners best? Impact of available partner- and content-related information on collaborative learning. In B. K. Smith, M. Borge, E. Mercier, & K. Y. Lim (Eds.), *Making a Difference: Prioritizing Equity and Access in CSCL, 12th International Conference on Computer Supported Collaborative Learning (CSCL) 2017* (Vol. 1, pp. 127–134). Philadelphia, PA: International Society of the Learning Sciences. Retrieved from https://repository.isls.org/bitstream/1/223/1/20.pdf

→ Study 3

Paper 3

Erkens, M., & Bodemer, D. (2019). Improving collaborative learning: Guiding knowledge exchange through the provision of information about learning partners and learning contents. *Computers & Education, 128*(1), 452–472. doi:10.1016/j.compedu.2018.10.009

→ Study 3

1. Research summary and background

Today's increasing digitalization requires many skills of instructors to design interactions in digital learning environments; at the same time, it creates innovative means of automating instructions for interaction support. This work explores the potential of text mining to support knowledge exchange. Knowledge exchange refers to social interactions in a group, which can promote learning, particularly when the communication is goal-oriented (Buder, 2017). As learning is desired not only in schools but also in workplaces and in private life, skills to collaborate and communicate with others in such goal-oriented manner are increasingly required. However, most people lack such skills and need guidance on how social interactions should be conducted (Kollar, Fischer, & Hesse, 2006; Weinberger, Stegmann, Fischer, & Mandl, 2007). To support these people, the knowledge of beneficial processes and instructions are brought together under the umbrella of collaborative learning (Lin, 2015). Collaborative learning offers different approaches that consider the conditions under which the knowledge exchange produces success. Moreover, it offers answers to the question on how to guide conversation partners to promote learning emanating from the exchange. Thus, collaborative learning is an important tool for instructors, offering a wide range of opportunities that maximize the benefits of knowledge exchange.

Against this background, different types of approaches have proven effective in supporting knowledge exchange, especially group formation and group awareness support. Group formation includes approaches that use a complementary or contrary distribution of cognitive characteristics among learning partners (Dillenbourg & Jermann, 2007). By distributing characteristics (e.g., knowledge or opinions) in this manner and providing guidance on how to use this distribution for learning, instructors can explicitly guide the learning processes of learning partners (Fischer, Kollar, Stegmann, & Wecker, 2013). Numerous studies (see Johnson, Johnson, & Stanne, 2000) have demonstrated the positive effect of this guidance approach on learning outcomes. In comparison, group awareness support includes approaches that merely suggest specific modes of thinking and behaving through the provision of information about the learning partners and learning content (Bodemer, 2011). Offering these types of information, which implicitly guides the knowledge exchange of people, also positively influences social interactions (see Bodemer, Janssen, & Schnaubert, 2018). Therefore, the use of both approaches is recommended for instructors to promote learning, the combination of which seems reasonable to enhance their effects (see *Guiding learners explicitly and implicitly* in Paper 1).

Both approaches require that instructors have cognitive information about the conversation partners. Hence, instructors initially need to collect this information and sometimes even transform it, and both tasks are effortful. The procurement of information in preparation for supporting collaborative learning does often not seem to meet the demand for efficiency. Computer support plays an important role in solving this problem because it can complement efficiency by automating the processing of information to some extent and thus facilitate instructional guidance. Text mining can be one means of facilitating group formation and group awareness support (see *Text mining as a basis for forming groups and representing cognitive information* in Paper 1). Text-mining methods can automatically transfer unstructured texts representing their authors' knowledge into a structured format (Miner et al., 2012), for example, into clusters of concepts interpretable as topics or text clustering-based distance values signifying the topical differences of texts. This cognitive information can be used for informing group formation or providing the people involved with visualizations for group awareness support. However, although text mining-based support might have substantial potential to reduce instructors' effort in instructional guidance, it has not yet been applied in the collection and transformation of information to eventually support knowledge exchange.

For implementing text-mining methods in knowledge exchange, instructional guidance needs to be integrated with sophisticated methods from computer science under consideration of the mechanisms underlying collaborative learning. For this integration, three problems need to be solved. (1) First, numerous text-mining approaches and methods are available. Therefore, the most suitable methods for collecting and converting cognitive information to inform group formation and facilitate group awareness support must be selected. (2) Second, as group formation and group awareness support have not yet been combined based on text-mining methods, the effect of using text mining-generated information on learning requires validation. (3) Third, text-mining methods can yield different types of information to be provided to learning partners, the individual effects of which have generally remained unexamined. Furthermore, the text-mining methods allow for the bottom-up generation of content-related information that has yet to be utilized for instructional guidance. Hence, research should further investigate the general mechanisms underlying knowledge exchange as well as mechanisms in the specific case of text mining-based support. In summary, the application of text-mining methods to support knowledge exchange depends on overcoming the aforementioned problems, which requires the analyses of text-mining methods and the processes associated with learning.

The objectives of this work are related to the preceding problems. Regarding problem (1), the aim is to test and compare text-mining methods in terms of their suitability for collecting and transforming cognitive information for didactic support. The result is a selection of text-mining methods, which allows for an integrated approach of group formation and group awareness support in one tool (see Paper 1). Concerning problem (2), the purpose is to validate the previously selected text-mining methods in the aspect of appropriateness for improving collaborative learning. The outcomes are suggestions for improving the tool design in terms of text-mining functions (see Paper 1). With regard to problem (3), the goal is to test the influence of providing different types of information that can result from the selected text-mining methods or other sources. Thus, the focus is on the general comprehension of learning processes resulting from the reception of various types of provided information but also on the goal of making more specific statements about text mining-generated cognitive information. The results are concrete proposals for designing text mining-based tools in particular or modifying existing tools for improving the guidance of knowledge exchange in general (see Papers 2 and 3). In summary, the objective of this work is to select the suitable text-mining methods for informing group formation and facilitating group awareness support, and hence examine the effect of providing cognitive information generated by these methods and comprehend the underlying mechanisms.

Three empirical studies have been conducted to achieve the aforementioned objectives. Study 1 compares different text-mining methods in terms of their suitability for collecting and transforming cognitive information and selects methods for informing group formation and facilitating group awareness support. Study 2 tests the selected methods in the field in terms of their effect on collaborative learning. Finally, Study 3 examines the influence of providing different types of cognitive information, i.e., information about learning partners and learning content, on the learning processes required for learning.

The subsequent sections of this work bring together the different foci of the three studies in an integrated theoretical background. In particular, these sections clarify the importance of knowledge exchange in collaborative learning by explaining the mechanisms that need to be triggered to promote learning (section 1.1.), how knowledge exchange can be supported by guidance (section 1.2.), and what role text-mining methods can play in the support of knowledge exchange (section 1.3.). Figure 1 (p. 4) illustrates the specific section that is in the focus of which study and how the individual studies are interrelated. The theoretical background concludes with the formulation of research questions. This work subsequently presents a comprehensive summary of the studies' results and a discussion of these findings, offering implications and recommendations for future research on (text mining-supported) knowledge exchange in collaborative learning.

1.1. Mechanisms of knowledge exchange in collaborative learning

Learning can be achieved through different mechanisms. These mechanisms can be explained on the assumptions that (1) certain processes promote positive learning outcomes, and (2) these processes are promoted by specific conditions that depend on the individual or can be induced by the situation (Wecker & Fischer, 2014). As the current work focuses on learning improvement through knowledge exchange, it underscores the mechanisms that improve cognitive learning outcomes and, as a first step, identifies the processes that are conducive for learning and under which conditions they transpire, therefore referring to collaborative learning. Collaborative learning, similar to knowledge exchange, relates to the social interactions of at least two individuals (Dillenbourg, 1999), but above all empowers people to achieve the goal of learning or developing a shared understanding through these social interactions (e.g., Dillenbourg, 1999; Roschelle & Teasley, 1995; Scardamalia & Bereiter, 1994). For this purpose, collaborative learning defines the processes and conditions that are conducive for achieving the aforementioned goals (Lipponen, 2002). The model of knowledge exchange in Figure 2 (p. 5) indicates from the perspective of an individual that cognitive processing, involving cognitive and metacognitive processes at an internal level, and communication, asking questions and giving explanations at an external level, are considered in collaborative learning (see Buder, 2017; Wecker & Fischer, 2014). This model also illustrates that the quality of the internal representations of own and other's cognitive characteristics, which are stored in memory (e.g., available content of knowledge or high or

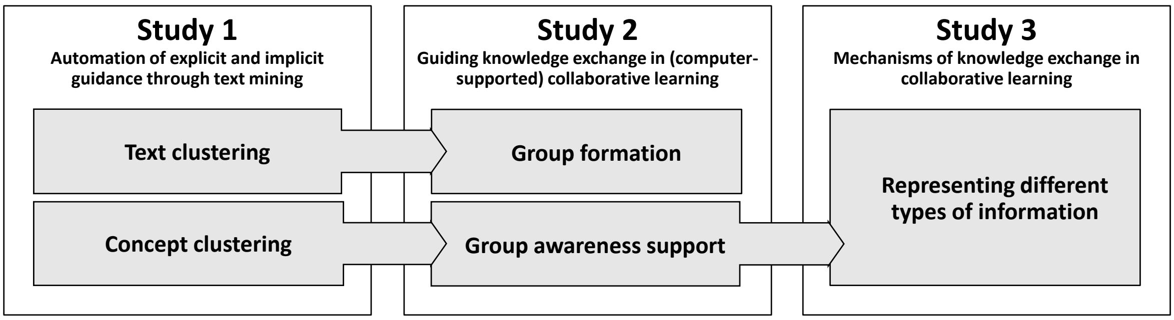

Figure 1. Overview of studies and their focus of research. Study 1 focuses on the selection of text-mining methods for text and concept clustering to automate instructional guidance (see section 1.3.). Study 2 explores the guidance effect of group formation and group awareness support (see section 1.2.) based on the selected text-mining methods. Study 3 examines the cognitive, metacognitive, and communication processes (see section 1.1.) triggered by representing information about learning partners and about learning content.

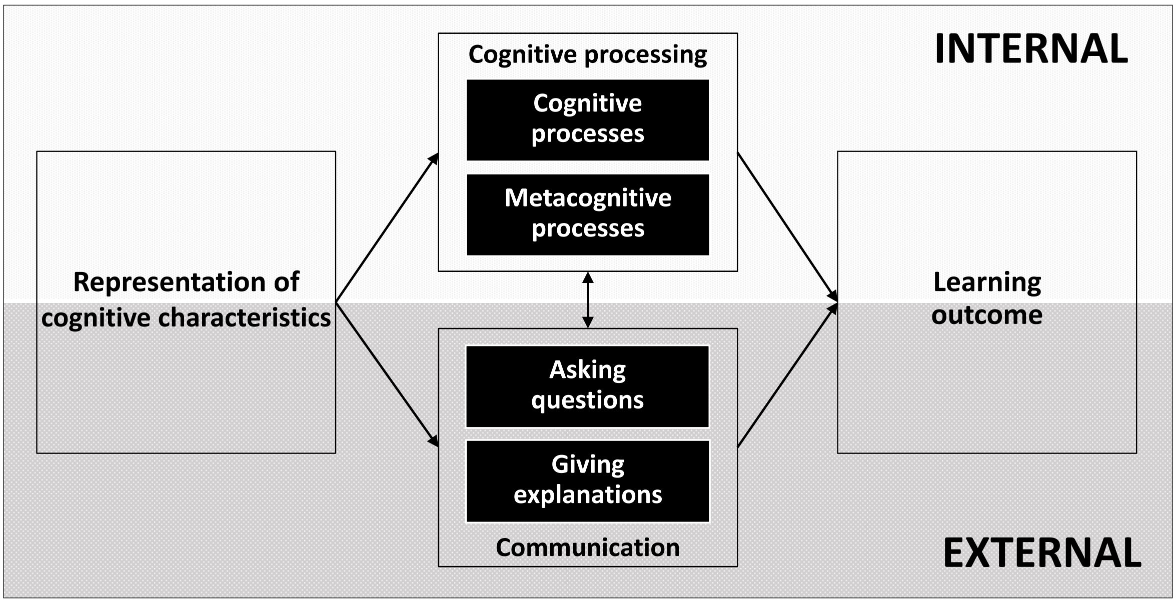

Figure 2. Knowledge exchange from an individual perspective. Specific cognitive and metacognitive processes (internal cognitive processing) and modes of asking and explaining (external communication) can improve learning outcomes; certain internal (e.g., low levels of own knowledge) and external representations of cognitive characteristics (e.g., specific content of questions and explanations of others) can trigger these processes.

low levels of one's own and other's knowledge), influence these processes (see Buder, 2017; Wecker & Fischer, 2014). In the same way, external representations, which are observable during the exchange (i.e., verbal contributions from the learning partners such as others' questions and explanations on content or their statements about high or low knowledge levels), influence the processes (see Buder, 2017; Wecker & Fischer, 2014). Understanding the interdependencies between conditions of representations, learning processes, and learning outcomes is fundamental to the current work. Hence, this section explains in more detail than the theoretical parts of the included papers the emergence of mechanisms.

1.1.1. Cognitive processing: conditions and processes conducive to learning

Cognitive processing involves cognitive and metacognitive processes that can promote learning under certain conditions. Cognitive processes refer to an individual's mental selection, organization, and integration of information into a coherent knowledge structure (Mayer, 2010; Moreno & Mayer, 2007). In knowledge exchange, an individual might attend to a concept from a conversation partner's explanation (selection), link it to another concept from existing knowledge (organization), and store it in memory (integration); by contrast, no cognitive processes would occur if the individual did not attend to the concept. Metacognitive processes pertain to the mental monitoring and control of cognitive processes

(Dunlosky & Metcalfe, 2009; Nelson & Narens, 1994). Monitoring covers processes that inform about cognitions and generates models stored in memory, whereas control involves the processes of initiating, continuing, or terminating cognitive processes (Nelson & Narens, 1990). For instance, the same individual from above might judge own knowledge on the concept as being insufficient (by monitoring) and further elaborate on the concept to verify the achieved misunderstanding (by controlling). Another conceivable idea is that the existing knowledge on the concept would be rated as sufficient so that no subsequent processes would be initiated. These examples demonstrate that both cognitive and metacognitive processes depend on the internal and external representations of cognitive characteristics, and they can both be consequently promoted by fulfilling some conditions.

Prior knowledge activation is a cognitive process that can promote learning. Prior knowledge refers to people's knowledge that is available before working on a certain learning task (Dochy, de Rijdt, & Dyck, 2002). This knowledge needs to be activated for learning (see Ausubel, 2000; Ausubel, Novak, & Hanesian, 1978; Mayer, 2010; Schneider, Körkel, & Weinert, 1989; Weinert & Helmke, 1998), which indicates bringing it from long-term memory to working memory (Mayer, 2010; Mayer, 2001). Once activated, prior knowledge in the working memory can be compared to perceived information and integrated with this information (Ausubel, 2000; Mayer, 2001; Mayer, 1979), or it can be reorganized for integration (Linn, 2005; Schnotz & Preuß, 1999) where learning occurs when these new schemata are successfully transferred from working memory to long-term memory (Mayer, 2001). Thus, the following conditions should be fulfilled: (1) prior knowledge about the learning content should be given, which is an individual prerequisite, and (2) prior knowledge about the learning content should be identified and activated.

Supporting prior knowledge activation requires drawing the learners' attention to relevant learning content. A learning situation that helps to fulfill the condition that learners know the specific contents of prior knowledge that are to be activated can consequently have a positive effect on their learning. For example, the instruction to reflect on what prior knowledge is important for fulfilling a task can enhance learning (Dochy, Segers, & Buehl, 1999). In addition to direct instructions, the advance provision of "higher level concepts" can facilitate prior knowledge activation (Ausubel, 2000). Higher level concepts are concepts at a higher level of generality; for example, learners who have never heard about the *t*-test from learning material might still have an idea about a teacher comparing the performance of two groups, where group comparison is a higher level concept (Gurlitt, Dummel, Schuster, & Nückles, 2011). The advance provision of higher level concepts to which learners can subsume new knowledge from the learning content improves learning (see Ausubel, 2000). These findings from individual learning suggest that knowledge exchange, in which the new information to be integrated (learning content) is a contribution of the learning partner, might profit from the external representations of what learning content is relevant for being activated, thereby restricting the communication space to relevant topics. For more information of the advantages of prior knowledge activation in the context of collaborative learning, see also *1.1.2. Prior knowledge activation in CSCL* in Paper 3 (Erkens & Bodemer, 2019).

Another cognitive process for promoting learning is cognitive elaboration. Cognitive elaboration refers to the mental processes required for the above mentioned integration of different pieces of knowledge into coherent statements (Stegmann, Wecker, Weinberger, & Fischer, 2012), which includes organizing, interconnecting, and integrating new elements of information as well as relating these elements to existing knowledge (Kalyuga, 2009). Thus, cognitive elaboration not only reveals whether new mental models (external representation) and existing knowledge (internal representation) can form coherent schemata but also determines whether existing contradictions need to be resolved; without coherent schemata, a transfer from working memory into long-term memory might not be executed (see Mayer, 2001). As the ones who cognitively elaborate on explanations, either for self-explanations or explanations to others, can improve their learning (Ploetzner, Dillenbourg, Preier, & Traum, 1999), their focus of elaboration should be on explaining the relevant aspects of the learning content. Furthermore, the ones who receive elaborated explanations can increase their learning (Crommelinck & Anseel, 2013). Above all, a reasonable occurrence is when learning partners elaborate on content that is new for the receivers or not comprehended by them. Receiving cognitively elaborated explanations on content referring to knowledge gaps or inconsistencies might simplify the integration of the explained concepts into own cognitive structures. Thus, in knowledge exchange, the following conditions should be fulfilled for learning: (1) learners' explanations in knowledge exchange should focus on learning content, and (2) learners should be aware of one's own and learning partners' knowledge to concentrate on differences.

Regarding cognitive elaboration, it is conducive for learning to make learning partners aware of and focus their attention on the content-related differences between them. A comprehensive analysis of research indicates that cognitive elaboration can reveal knowledge gaps or inconsistencies of own concepts and concepts from the learning material, which can further trigger strategies for filling these gaps or resolving the inconsistencies

(Webb, 1989). In knowledge exchange, the learning partner is the information resource to fill gaps and simultaneously discover and resolve inconsistencies; hence, learning is only possible when the learning partner can provide the necessary help (Webb, 1989). For this purpose, learning partners should be aware of their differences to be efficient in their knowledge exchange. Thus, external representations drawing attention to content-related differences between learning partners can improve learning.

A metacognitive process that can promote learning is the accurate judgement about cognitive characteristics. Judging cognitions is a monitoring function (Nelson & Narens, 1994) that comprises estimates not only about one's own but also others' cognitions (Buder, 2017). The metacognitive judgements can result from social comparison (Salonen, Vauras, & Efklides, 2005) or from inferences based on available information (Nelson, 1996). Furthermore, they can be based on analytic processes or on non-analytical, often heuristic processes (Kahneman, 2003), the latter causing metacognitive judgements to be inaccurate (Koriat, 2007). The execution of strategies depends on these judgements regardless of whether they are used directly or are saved as models of own and others' knowledge in memory and only later accessed (see Efklides, 2008); hence, the learning outcome also depends on whether these judgements are accurate. Thus, in knowledge exchange, the following conditions should be fulfilled for learning: (1) learners should know about one's own and learning partners' knowledge, and (2) these judgements should be accurate.

To ensure accurate judgements, valid information about the learning partners should be available to avoid cognitive biases resulting from non-analytical processes. Cognitive biases are judgements that systematically deviate from the norm or rationality (see Baron, Voss, Perkins, & Segal, 1991). For instance, people without sufficient expertise can have the tendency to overestimate their own and underestimate their more knowledgeable conversation partners' abilities (Kruger & Dunning, 1999). Another example is the tendency of people to erroneously assume that other people have the same abilities as they have (Nickerson, 1999). Both examples indicate that without the correction of inaccurate judgements, wrong metacognitive strategies might be selected. Thus, knowledge exchange, the success of which depends on the accurate assessment of the level of all the participants' knowledge to evoke appropriate strategies, might profit from external representations signifying this accurate assessment of learning partners. However, as cognitive and metacognitive processes can affect behavior, communication can influence cognitive processing, as described in the next section.

1.1.2. Communication: conditions and processes conducive to learning

In addition to cognitive processing, communicating can promote learning, when asking questions and giving explanations are performed in a specific manner. Asking questions denotes the externalization of statements of inquiry (Callender, 2012). In knowledge exchange, an individual might not know a concept from a learning task (e.g., rules of three in mathematics) and ask some learning partners about the concept. One learning partner might provide a helpful answer, while the other one might offer a wrong explanation because he intends to help but does not know much more about the concept himself. Explaining to others connotes externalizing assumptions that would otherwise remain tacit (Ploetzner et al., 1999). For instance, the aforementioned knowledgeable learning partner might respond to the question by explaining the concept of the rule of three with a calculation example because he knows about the beginner ship of the learner who is asking; otherwise, he might simply tell the solution of the task without further explanation as he considers the receiver to be equal in knowledge. These examples illustrate that both asking questions and giving explanations depend on how the knowledge is distributed across the learners in a learning group and on the internal or external representations about this distribution that can steer communication.

Asking questions can promote learning when the questions address knowledge gaps and are directed to more knowledgeable learning partners. Asking questions is associated with learning because it can trigger cognitive processing. On the one hand, people have to examine whether the contents have been understood when they generate questions, thereby performing better monitoring (Palincsar & Brown, 1984). On the other hand, asking questions is associated with using the learning partners as a resource to receive the required information (Weinberger & Fischer, 2006). As questions are followed by some type of answer (Callender, 2012), learners have the opportunity to learn from answers (external representations), only if they get help from the learning partner and the explanation is elaborated (Webb, 1989). For instance, King (1994) suggested that learners who extensively ask their learning partners for elaborated explanations (and obtain answers due to a collaboration) learn more successfully than learners who ask fewer questions during collaboration. Thus, in knowledge exchange, the following two conditions should be fulfilled for learning: (1) questions should address topics about which the learner himself has knowledge gaps or comprehension problems, and (2) they should be directed to learning partners who can provide explanations.

Explaining content to less knowledgeable learning partners can promote learning when the learner is aware of the learning partner's knowledge. Explaining to others can promote learning due to the triggering of cognitive processing by such behavior: explaining to others can reveal knowledge gaps that need to be filled or inconsistencies that explainers might attempt to resolve (Webb, 1989), for instance, by elaborating on examples (Stark, Mandl, Gruber, & Renkl, 2002). As in this example, especially explainers who cognitively elaborate their statements can learn from explaining because the amount of learning is related to the cognitive activities necessary for constructing and externalizing explanations (see Webb, 1989). The level of elaboration depends on how the explainer assesses the knowledge of the questioner, as explainers adjust their communication toward their recipients, which is called audience design (see Clark & Murphy, 1982; Lockridge & Brennan, 2002; Schober & Brennan, 2003), for which, however, judgements about one's own and others' knowledge should be accurate. Thus, in knowledge exchange, the following two conditions should be fulfilled for learning: (1) explanations should address topics about which the learning partner has knowledge gaps or comprehension problems, and (2) they should be cognitively elaborated.

Content-related differences among learning partners should be given and be made aware of to trigger questions and explanations that are beneficial for learning. If learners adapt their communication to own and others' knowledge gaps or inconsistent ideas, they could profit from differences in a Piagetian (e.g., Piaget, 1977; Piaget, 1959) or a Vygotskian (e.g., Vygotsky, 1978) sense. From the Piagetian perspective (e.g., Piaget, 1977; Piaget, 1959), socio-cognitive conflicts are beneficial for learning meaning that the learning partners have different perspectives about the content or varied answers to a task (e.g., Ames & Murray, 1982; Doise & Hanselmann, 1991; Doise & Mugny, 1984). In the case of such conflicts, a disequilibrium arises in the individuals, which can be remedied by restoring the equilibrium (Piaget, 1959). Therefore, these individuals become active, seek new information, and engage in discussions (Johnson & Johnson, 2009). To be more precise, the individuals communicate and demand reasons, explanations, or justifications, thereby enabling them to recognize and fill knowledge gaps, identify and resolve inconsistencies, and construct more elaborate conceptualizations (Fawcett & Garton, 2005). The exchange of knowledge, in which questions and explanations should be based on the learning partners' differences, might profit from diverse opinions between learners and external representations of the contradictory perspectives.

Furthermore, differences in knowledge levels can be of relevance. From the Vygotskian (e.g., Vygotsky, 1978) perspective, variances in the learning partners' expertise levels are beneficial for learning, as learning partners can use this information to evaluate themselves (Bandura & Jourden, 1991; Festinger, 1954). Uncovered differences can contribute to recognizing the individual's zone of proximal development, which is known as the difference between what individuals can accomplish on their own and what they can accomplish with the support of a more "expert" partner (Fawcett & Garton, 2005). Being aware of such differences can also make less knowledgeable partners better match their requests for information to partners with more expertise (Neugebauer, Ray, & Sassenberg, 2016) and allow for more knowledgeable partners to adapt their communication (Ray, Neugebauer, Sassenberg, Buder, & Hesse, 2013). However, the exchange should transpire reciprocally so that both benefit from each other; otherwise, the expert could withhold information (Neugebauer et al., 2016). If reciprocity occurs, experts externalize their knowledge on content, enable the less knowledgeable ones to fill knowledge gaps, correct misconceptions, and develop or strengthen connections between new information and given knowledge (Fawcett & Garton, 2005). Thus, knowledge exchange, in which questions and explanations should be based on the learning partners' differences, might profit from the complementary knowledge of learning partners and external representations of the differing fields of expertise.

1.1.3. Summary

To ensure that knowledge exchange promotes learning, cognitive processing and communication can be positively influenced by making specific information available or more salient than the other. On the one hand, research has reported that cognitive processing promotes learning if prior knowledge on task-relevant content is activated and if the levels of own and others' knowledge are aware and accurate, because fulfilling these conditions facilitates cognitive and metacognitive processes. On the other hand, communication promotes learning if the differences between the learning partners' opinions and status of expertise are uncovered because this factor prompts learning partners to make a better choice of topics and offer better explanations. To make sure that learners perform such beneficial cognitive, metacognitive, and communication processes, they need support in perceiving or requesting the required information. Collaborative learning, in addition to defining useful collaborative processes, provides such support by affecting internal and external representations in such a way that certain learning-promoting processes occur.

1.2. Guiding knowledge exchange in (computer-supported) collaborative learning

Aside from answering the question of what individual conditions must be fulfilled for processes to occur that induce learning in knowledge exchange, collaborative learning provides suggestions on how conditions can be changed by instructions. Instructions are a core aspect of collaborative learning (e.g., Gokhale, 1995; Jacobs, Power, & Loh, 2016), and they refer to manipulations of the learning environment to guide learners in executing the beneficial processes described in the preceding section (see Mayer, 2001; Reimann, 2018; Romero & Lambropoulos, 2017; Wecker & Fischer, 2014). The extent to which learners are guided by instructions can considerably vary and depends on whether a more explicit or implicit approach is selected (Hesse, 2007; Reimann, 2018; Romero & Lambropoulos, 2017; Scardamalia & Bereiter, 2014). Explicit guidance denotes direct instructions to execute certain processes (e.g., Fischer et al., 2013; Weinberger, Stegmann, & Fischer, 2010), which indicates that in this approach, external representations are addressed by shaping the contributions of learning partners (e.g., through group formation); by contrast, implicit guidance connotes proposing useful processes by promoting awareness (e.g., of the learning partners' cognitive characteristics, Bodemer, 2011). This notion underscores the value of providing additional external representations that contain information about the learning content and the learning partners. Figure 3 (p. 13) depicts how group formation and group awareness support can affect learning processes and outcomes and how they can be combined. Both approaches are described in more detail in the following sections.

1.2.1. Explicit guidance of learners through group formation

Explicit guidance addresses learning in knowledge exchange by instructing learners with whom and how they should perform cognitive processing and communication. As previously described, explicit guidance refers to giving the learners a more detailed specification of the collaborative process, for example, in the form of collaboration scripts (see Fischer et al., 2013). Collaboration scripts are sets of mostly textual or graphical representations of collaborative practice (Fischer et al., 2013) that are similar to movie scripts, and these collaboration scripts specify how group members have to interact (Dillenbourg, 2002). A powerful mechanism of scripts is the formation of groups, which denotes that instructors distribute learners' individual characteristics across groups in a specific manner (Kobbe et al., 2007). In most cases, the distribution targets the heterogeneity of the learners' individual

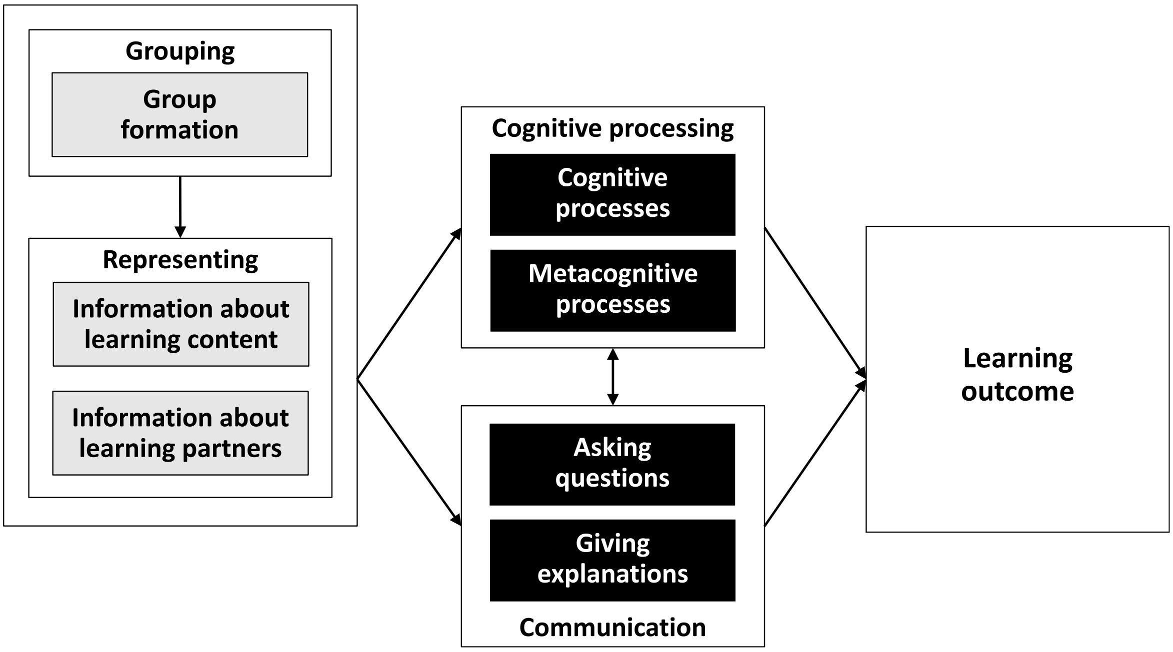

Figure 3. Combination of guidance approaches to promote learning. Collaborative learning can positively influence the learning situation through group formation and the representation of information that helps to improve the learning group's awareness.

characteristics by using one of these two variants: knowledge-complementary distribution or conflictual distribution (Dillenbourg & Jermann, 2007), which use the difference-based mechanisms described previously (see section 1.1.2. and also *Guiding learners explicitly and implicitly* in Paper 1, Erkens, Bodemer, & Hoppe, 2016). A knowledge-complementary distribution involves the formation of pairs or groups of learners with complementary knowledge (e.g., in the UniverSanté script, Berger et al., 2001), or the provision of complementary information to teammates (e.g., in the Jigsaw script, Aronson, Blaney, Stephan, Sikes, & Snapp, 1978). On the contrary, a conflictual distribution of participants entails the establishment of groups based on the learners' conflicting views (e.g., in the ArgueGraph script, Dillenbourg & Jermann, 2007), or the provision of conflictual information to teammates (e.g., in the structured controversy script, Johnson & Johnson, 1979). These examples illustrate that forming groups with heterogeneity-based distributions has already been used in many variants.

In addition to the information on the distribution and with whom to collaborate, collaboration scripts provide instructions on how learners should interact with their assigned learning partners, ensuring the appearance of the desired learning processes. Instructions in collaboration scripts can address cognitive, metacognitive, or social processes (Mäkitalo-Siegl & Kollar, 2012), the latter including communication. To explain the manner of addressing the processes and shaping the external representations, the complementarity-based Jigsaw script (Aronson et al., 1978) is selected. In this case, cognitive processes are addressed by instructing the learning partners to exchange about their different areas of expertise to enhance the integration of the content of their own field of expertise with the contributed content from others' expert areas. Moreover, metacognition is tackled by instructing learners to identify knowledge gaps or inconsistencies in expert groups. Finally, communication is addressed by instructing the learners to take turns so that all the members in a group benefit from the communication. Thus, to foster the desired cognitive processing and communication, instructors can use heterogeneity-based group formations, which can be applied by inducing or detecting divergent knowledge or opinions.

When applying the mechanism of group formation, learners might be more engaged when they are not guided too extensively. Numerous studies have revealed the positive effect of group formation based on the learners' induced heterogeneity on learning outcomes (see Johnson et al., 2000). However, one criticism of such induced roles is that they prompt learners to merely play interactions similar to a teacher–learner game and hence cause learners to miss engagement (Dillenbourg, 2002), for example, when learners simply recite an opinion from a text instead of representing their own opinion with enthusiasm. Another criticism is that an extremely strong explicit guidance, which might be the case with induced knowledge but also with to too strong specification of the communication, can diminish engagement due to less intrinsic motivation (Hesse, 2007). Thus, engagement might be fostered if group formation is (1) based on the learners' authentic cognitive characteristics and (2) combined with less stringent communication guidelines, which is the case with group awareness support. For an explanation why both approaches should be combined, see also *Guiding learners explicitly and implicitly* in Paper 1 (Erkens et al., 2016).

However, grouping learners based on given knowledge-complementarity or conflictual opinions might involve some effort on the part of instructors, who must initially collect information to form groups. For instance, to apply the UniverSanté script (Berger et al., 2001) that uses knowledge differences by allowing students from various countries to discuss medical cases (Dillenbourg & Jermann, 2007), instructors must first identify the differences in medical education between different countries to decide what countries to group and to successfully apply the mechanism. In the ArgueGraph script (Dillenbourg & Jermann, 2007), which visualizes the opinions of learning partners by their positioning in a coordinate system, learner inputs in a questionnaire are required to illustrate their opinion on two axes. Therefore, instructors must initially design a sophisticated questionnaire that can map the two dimensions of the coordinate system and the extrema of the two axes. As this procedure costs time and effort, supporting the instructors in this work is reasonable (see also *Guiding learners explicitly and implicitly* in Paper 1, Erkens et al., 2016). Overall, to foster learners' engagement while they are interacting, instructors can especially use heterogeneous group formations based on given knowledge differences or opinions. Nevertheless, this approach entails increased effort in preparing the grouping, and it should therefore be supported.

1.2.2. Implicit guidance of learners through group awareness support

Similar to explicit guidance, implicit guidance addresses learning processes in knowledge exchange by representing information. However, in this case, the information provided merely stimulates the learners' certain modes of thinking, communicating, and behaving instead of directly instructing them to perform specific activities (Bodemer, 2011). This stimulation is commonly induced by cognitive group awareness tools (Bodemer & Dehler, 2011; Janssen & Bodemer, 2013) that have two specific features to support knowledge exchange in terms of the mechanisms mentioned previously. One feature is that these tools inform learners about the cognitive characteristics of their learning partner(s), for example, the latter's knowledge levels (Sangin, Molinari, Nüssli, & Dillenbourg, 2011). When the provided information refers to all the members of a learning group to facilitate comparisons, this feature also includes the sub-feature of providing self-related information (Bodemer et al., 2018). The second feature is that the tools provide information about the learning content, as they refer information about learning partners to more or less specified content (Bodemer et al., 2018). Therefore, these tools inform about often preselected parts of the learning material (Bodemer et al., 2018), for example, about learning modules (Sangin et al., 2011). To provide learners with both types of information, cognitive group awareness tools commonly use visualizations (Buder & Bodemer, 2008) in which the information about learning partners is often visualized by graphics such as bar charts, which illustrate the extent of knowledge or comprehension of learning partners (e.g., Sangin, Molinari, Nüssli, & Dillenbourg, 2008; Sangin et al., 2011). Information about the content is usually visualized as text, for example, by listing the topics of the learning material (Dehler, Bodemer, Buder, & Hesse, 2009; Dehler, Bodemer, Buder, & Hesse, 2011; Dehler Zufferey, Bodemer, Buder, & Hesse, 2011). Based on these features, the guidance effect arises because the additional external representation can highlight certain aspects of the situation, subsequently triggering activities.

Providing the aforementioned types of information fulfills several functions to enhance learning. Some important functions are (see Bodemer & Scholvien 2014; Bodemer et al., 2018; Dillenbourg & Bétrancourt 2006): to facilitate partner modeling, cue essential information about the learning content and constrain the content of communication (see also *Guiding learners explicitly and implicitly* in Paper 1, Erkens et al., 2016). Accordingly, cognitive, metacognitive, and communication processes are promoted by the provided information. Cognitive processes, for instance, can be successfully addressed by providing learners with the information that they have more knowledge on a topic than the learning partner, which induces them to better verbally and thus cognitively elaborate the content explained to their learning partner (e.g., Dehler Zufferey et al., 2011). Metacognition can be tackled by providing the self-related information that learners have little knowledge on a topic, which prompts them to ask questions on the content (Dehler et al., 2011). Communication processes can be addressed by informing learners about cognitive differences in content so that learning groups discuss those topics where differences emerge (e.g., Bodemer, 2011; Dehler et al., 2011; Dehler Zufferey et al., 2011). Thus, external representations provided to learners to support cognitive group awareness can promote the performance of cognitive, metacognitive, and communication processes. Dependent on its type, information about learning partners or information about learning content, the information provided might fulfill different functions with regard to beneficial learning processes during knowledge exchange, but this has not yet been investigated further. For a possible breakdown of supported mechanisms by information type, see *1.2. Functions of cognitive group awareness tools in CSCL* in Paper 3 (Erkens & Bodemer, 2019).

To provide both types of information accurately, their collection might better be based on objective measures than on learners' self-assessment. Prior to visualizing information that evokes the aforementioned learning processes, information about learners' knowledge or about their views needs to be collected and transformed (Bodemer & Buder, 2006). The collection of input data can be based on different instruments, which might require students' subjective evaluation, for example, self-assessment of knowledge (e.g., Dehler et al., 2011), or it might be based on objective indicators such as the results of knowledge tests (e.g., Sangin et al., 2011). Self-assessed cognitive characteristics might be associated with the problem of being biased (see section 1.1.1.); hence, selecting objective measures seems to be a better approach. However, based on either self-assessment or objective instruments, instructors have to design tools for collecting data, such as opinion polls and knowledge tests that match the learning material, where the objective survey is likely to be more timeconsuming (see also *Guiding learners explicitly and implicitly* in Paper 1, Erkens et al., 2016). Thus, to improve learners' learning processes, instructors might better support group awareness by objectively collecting information about learning partners and learning content. However, this approach entails increased effort, and it should therefore be supported.

Group formation and group awareness support can benefit from being informed by the automated analysis of learner-generated text. Although instructors' effort is higher if group formation is based on given (instead of induced) differences and group awareness support is based on objectively captured (instead of self-assessment-based) measures, it is probably also more conducive to learning. However, there might also be a possibility to facilitate learning and still keep the instructors' effort low: automating the analysis of learner-generated text. The production of texts (e.g., essays) is part of learners' everyday school life, digitalization eases the texts' accessibility, and they represent an authentic picture of which topics learners know about and how much they know about them (see also *1.1.1. Cognitive group awareness in CSCL* in Paper 3, Erkens & Bodemer, 2019). Furthermore, they might be an appropriate source to better initialize the activation of prior knowledge (see *1.1.2. Prior knowledge activation in CSCL* in Paper 3, Erkens & Bodemer, 2019). However, as the task to read learners' texts is complex and never entirely objective, the automation of this analysis can help, for which learning analytics is suitable. Learning analytics is a new form of assessment instrument for supporting educational practice (Knight & Buckingham Shum, 2017), which offers various automated approaches for collecting, transforming, and reporting machine-readable (big) data about learners, amongst others learner-generated texts (see Kovanović, Joksimović, Gašević, Hatala, & Siemens, 2017), to improve learning processes and learning environments (see Ferguson, 2012). Thus, the automated analysis of learner-generated texts offers the possibility to collect information about the learners' cognitive characteristics (due to their texts representing knowledge) not only efficiently (due to automation) but also objectively (due to the use of algorithms)**.** After the summary of this section, it is further explained how such analysis of learner-generated text can be particularly suitable for group formation and group awareness support (see also *Text mining as a basis for forming groups and representing cognitive information* in Paper 1 in Erkens et al., 2016).

1.2.4. Summary

For instructors, a promising means of guiding learners' knowledge exchange is a combination of group formation and group awareness support that can be simplified by the automated analysis of learner-generated content. In knowledge exchange, guidance is required to promote beneficial learning processes, where learners can be guided to varying degrees depending on the approach selected, explicit or implicit guidance. Explicit guidance often involves group formation to distribute cognitive characteristics across learning groups as a basis for instructed learning activities being successful. On the contrary, implicit guidance leaves more room for learners' self-regulation, as representing cognitive information does not directly instruct but suggests certain activities with regard to cognitive processing and communication. The combination of grouping and representing seems reasonable to trigger or maintain beneficial learning processes, whereby the heterogeneity-based grouping is the optimal basis for the representations of information to promote the reciprocal exchanges of co-learners. However, it is advisable that the cognitive information for group formation and group awareness support is collected objectively and based on given characteristics, which involves some effort for the instructors. To relieve them of this effort, the automated analysis of learner-generated text using learning analytics is reasonable, which can be applied with the aid of text mining.

1.3. Automation of guidance through text mining

To automate guidance, an appropriate approach of learning analytics is text mining. Text mining is an established computational technique for automating the analysis of learner-generated texts (see Hoppe, 2017). It is characterized by offering efficient methods for bringing digital text, which is usually given in an unstructured form, into a structured format (Miner et al., 2012). In order to combine this automated transformation with group formation and group awareness support, Paper 1 (Erkens et al., 2016) proposes an integrated tool design that offers answers on how guidance can be automated by the support of text-mining methods. Figure 4 (p. 19) illustrates the functions of this tool: In a first step, relevant concepts are identified, generating a preprocessed text corpus. This preprocessed text corpus is the basis for further analyses, whereby the assessment of text closeness can assist group formation and the evaluation of concept closeness can aid group awareness support. In the following sections, the functions of text mining-based grouping and representing are explained in more detail, as they are also described in Paper 1 (see *Specification of the functions of the GRT* in Erkens et al., 2016). In addition, the subsequent sections elaborate the advantages that these text mining-based functions can bring in terms of applying the previously outlined mechanisms.

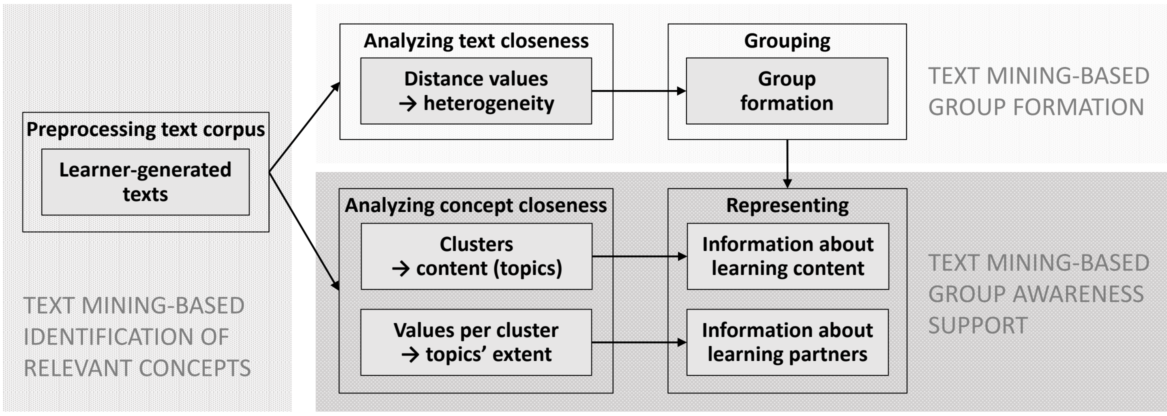

Figure 4. Schematic representation of text mining-based grouping and representing (Erkens et al., 2016). The preprocessing identifies relevant concepts and informs the analysis of text and concept closeness. The assessment of text closeness further informs the grouping by distance values. The evaluation of concept clusters facilitates the representation of information by providing concept clusters (information about learning content) and generating values per cluster (information about each learning partner to what extent the topics are given in a text).

1.3.1. Text mining-based identification of relevant concepts

Before applying text-mining methods to inform group formation and facilitate group awareness support, the preparation of the text corpus is necessary for the specification of content. This preparation usually includes the following steps: deleting stopwords, identifying collocated words, and stemming (Miner et al., 2012). Stopwords are superfluous words (e.g., "the" and "and") that appear in nearly every text and thus should be removed because they do not have an effect on the clustering results (Miner et al., 2012). Collocated words are groups of words that frequently occur together and represent a single idea, their detection differing from clustering as a collocation consists of a group of words that appear consecutively (Miner et al., 2012) and should be combined; for example, the collocated terms "climate" and "change" should be combined into the concept "climate change" for further analysis. Stemming pertains to the reduction of words to their root stem (Miner et al., 2012), which, for instance, would cause both the verb "warmed" and the noun "warming" to result in the concept "warm". This automated preprocessing is the basis for further steps to determine the relevance of content.

Subsequent to the above points, the relevance of the remaining concepts can be determined by a thesaurus and frequency measures. A previous argument is that instructors should help learners to identify content that is relevant to knowledge exchange (see sections 1.1.1. and 1.1.2.). This identification can not only promote the activation of prior knowledge but also the recognition of topics on which the partner has more knowledge, or has knowledge when own knowledge is missing. Thus, relevant content can be defined as such; that is, content that (1) represents technical terms (for referencing the learning topic), (2) is contained in the knowledge of all the involved learning partners (for recognizing those individuals who more or less know the same topic), or (3) is merely presented by some authors but more extensively (for identifying specialist knowledge where others' knowledge is missing). Regarding (1), technical terms can be identified and unified through a thesaurus. Concerning (2), the frequency of concept occurrences in texts can be counted as the more frequently preprocessed concepts appear, the more relevant they might be for the entire group (see Daems, Erkens, Malzahn, & Hoppe, 2014; Erkens, Daems, & Hoppe, 2014). With regard to (3), tf/idf values can help to identify specialist concepts. For more information on these measures, see also *Specification of the functions of the GRT* in Paper 1 (Erkens et al., 2016). In the current work, their application is the basis for obtaining more interpretable results from subsequent analyses.

1.3.2. Text mining-based group formation

If instructors intend to group learners according to their cognitive heterogeneity, their effort can be reduced through a text mining-based determination of the differences in learner-generated texts. The automated grouping of learners in the educational context, especially for discovering the skill models of learning, has often been based on the learners' interaction with the learning content (e.g., Liu & Koedinger, 2015; Nugent, Ayers, & Dean, 2009; Perera, Kay, Koprinska, Yacef, & Zaïane, 2009) but not on learner-generated content. To guide the exchange of knowledge, however, forming groups of learners on the basis of learner-generated text seems useful because comparing the text content of different learners might reveal their cognitive heterogeneity. The assumption is that the high values of text differences might be an indicator of heterogeneous knowledge. An example from Paper 1 (Erkens et al., 2016) illustrates this point. The example case involves one student who writes about the environmental effects of the coal-based generation of energy in comparison to nuclear energy. Meanwhile, another student writes about wind energy and its advantages in comparison to nuclear energy. Although both students write about energy, their text differences are likely to be high due to their varied topics with regard to which they can complement each other. Moreover, the heterogeneous knowledge may even substantiate divergent opinions. The first student may be inclined to accept nuclear power plants with all their risks because he knows about their advantages over coal-based energy generation; by contrast, the other student may probably be more skeptical toward nuclear energy plants. However, instructors need to read the texts of all the learners to locate the aforementioned differences; text mining can automate this work. Thus, text mining seems suitable for identifying the given differences between learners' cognitive characteristics. This potential raises the question of how to transform information from the texts into values representing cognitive heterogeneity and how to group learners based thereon.

To help the instructors' decision on whom to group, the analysis of text closeness can identify pairs of learners with significant cognitive differences. In text mining, the analysis of text closeness pursues the goal of understanding how documents are related to one another (Miner et al., 2012). For this purpose, the text corpus is usually converted into a vector space, with each vector representing a text and containing the characteristics of all the concepts of the corpus, for example, the concepts' occurrence frequency in each text (Hotho, Nürnberger, & Paaß, 2005). Based on such vector space models (VSMs), various distance metrics such as the well-known Euclidian distance can be generated between the vectors that form the basis for text clustering (Miner et al., 2012). Such distance metrics provide a structured format of the distance between vectors, which indicates that they quantify the semantic variance between the texts of two different authors or between different texts of one author. The higher the value, the higher the heterogeneity of the content. Thus, text-mining methods can transform text to generate distance values that—if the values are higher than average values—might be suitable for grouping learners. However, the most appropriate method for determining cognitive heterogeneity for this grouping function remains to be clarified.

Once a suitable method has been identified to fulfill this function, the text mining-based group formation decision can be used as a basis for further support. The heterogeneous grouping ensures that differences exist (one knows more/one knows less/one knows something else than the learning partner) which, when combined with the support of awareness of these differences in a Piagetian or Vygotskian sense (see section 1.1.2.), can help learners to focus on certain topics that are related to these differences. Hence, the assumption is that grouping forms the basis for achieving an effect on learning with the additional visualization of differences, whereas the effect would be less without such additional awareness support. This premise is illustrated in Figure 4 (p. 19) through the arrow from grouping to representing.

1.3.3. Text mining-based group awareness support

If instructors intend to support learners' group awareness, their effort can be reduced through the text mining-based specification of content. In the field of learning analytics, text-mining methods have already been applied to discover the knowledge of learners. For instance, Sherin (2012) has applied text-mining methods to identify topics (i.e., lists of clustered concepts) and specify therewith the students' knowledge content. Furthermore, text-mining methods have been successfully used for analyzing the development of such topics in texts over time, and the benefit of feeding back this information to students in terms of their writing at the collaborative level has been highlighted (Southavilay, Yacef, Reimann, & Calvo, 2013). Both the text mining-based transformation of texts into topics and the provision of the latter can facilitate the instructors' work. However, in the two aforementioned examples, different text-mining methods have been utilized for transforming the content from text into topics, and deciding how the transformation should optimally occur for the purpose of group awareness support would be necessary. Text-mining methods can therefore identify the relevant topics of a group of learners as a basis for feeding back this specified information about the content to the learners to implicitly guide them. Nevertheless, this potential raises the question on how to transform information from text with text-mining methods.

The analysis of concept closeness can help instructors in terms of specifying the content of written texts. Such analysis pursues the goal of understanding the meaning of texts (Miner et al., 2012). For this purpose, concepts are assigned to clusters (Aggarwal & Zhai, 2012). Depending on the selected algorithm, such clustering is called concept extraction or topic modeling (Miner et al., 2012). Concept extraction indicates whether concepts are used in the same semantic context (co-occurrence), where the context may be defined by sentences, paragraphs, or texts, and the clustering is executed with VSM-based similarity metrics (Aggarwal & Zhai, 2012). Topic modelling (e.g., based on the probabilistic topic model (PTM) called latent Dirichlet allocation (LDA)) is used for clustering concepts into topics, whereby concepts can probabilistically appear in multiple topics (Blei, 2012). In either manner, the outputs resulting from these different methods are lists of concepts for each cluster representing the semantic meaning of a text corpus and are interpretable as topics to specify the content. As mentioned previously, different text-mining methods have been used thus far to generate such lists in an educational context. Sherin (2012) utilized concept extraction based on a hierarchical clustering of vectors in a VSM, whereas Southavilay et al. (2013) used PTM in the form of the aforementioned LDA. However, these two instances of applied methods from the two areas of concept extraction and topic modelling have different characteristics that require a comparison. To conclude, several text-mining methods can automatically transform learner-generated texts into concept clusters, but they still need to be compared to ascertain their suitability for fulfilling the function of accurately specifying the content that is represented in the texts.

Once an appropriate method has been identified to fulfill this function, the text mining-generated topics or concept lists can be represented for group awareness support. The text mining-based analysis ensures that the concepts and/or topics provided (higher level concepts based on the interpretation of concept clusters; for example listings, see Erkens et al., 2016, and Erkens & Bodemer, 2019) refer to the learners' given knowledge, which is a bottom-up approach. Visualizing bottom-up concepts denotes the provision of the external representations of cognitive characteristics that are inherently related to internal representations; hence, this approach should improve cognitive processing and communication in knowledge exchange. On the one hand, the provision of such individualized information about content might improve the chances for co-learners' prior knowledge activation and thus the integration of new knowledge (condition for better cognitive processes, see section 1.1.1.). Furthermore, it can render the visibility of unfamiliar content (condition for better metacognitive processes, section 1.1.1.) because the provided information about the content also refers to the learning partner's knowledge, which represents what the learning partner can teach the learner verbally during the collaboration and can thus be called "information about learning content" (see middle row in Figure 4, p. 19). On the other hand, providing bottom-up information about such learning content might improve the communication (condition for better questions and explanations, see section 1.1.2.) because co-learners' tackling of more task-relevant topics in their conversation due to the better prior knowledge activation seems reasonable (see also section 1.1.1.).

In addition to facilitating the provision of text mining-generated information about the learning content, text-mining methods can reduce the instructors' effort to provide information about learning partners. Thus far, learner-generated text has been chiefly assessed in learning analytics to describe the quality of content by scoring the writing skills in terms of the cohesiveness of essays (e.g., Crossley, Allen, Snow, & McNamara, 2015; Dufty, Graesser, Louwerse, & McNamara, 2006; Foltz & Rosenstein, 2015; Tansomboon, Gerard, Vitale, & Linn, 2017). However, another reasonable approach is the identification of how much a learner has written about each topic in a text to describe a learner's expertise on the content. This information can indicate the individual learners' knowledge level for every specified topic, which can be reported back to them. Therefore, learner-generated texts need to be transformed into values per cluster to provide a basis for visualizing this information—self-related levels of knowledge, learning partner's levels of knowledge, knowledge levels in comparison to each other—to implicitly guide the learners.

With regard to offering this information about the learning partners, the text mining-based analysis of concepts can further help instructors to identify how much learners in a group wrote about the identified topics. Along with concept clusters, the evaluation of concept closeness can provide values representing how strongly each cluster is exemplified in a single text or rather how much an author wrote about a topic. This cognitive characteristic about learning partners is quantified by values representing how often specific concepts from their texts have been assigned to particular clusters and might indicate to a certain extent their knowledge levels (see bottom row in Figure 4, p. 19). This function therefore eases the instructors' generation of references to learners' self-related levels of knowledge, their learning partners' levels of knowledge, or special constellations of both. However, the most appropriate text-mining method for this purpose still needs clarification.

Once an appropriate method has been determined to fulfill this function, the topic extents can be used for group awareness support. Based on the aforementioned analysis, these values can be utilized by the instructors to (comparably) represent the learners' self-related topic extents, their learning partners' topic extents, or special constellations of both by visualizing the extents in bar charts and feeding it back to the co-learners (see the example bar charts in *Specification of the functions of the GRT* in Paper 1, Erkens et al., 2016). As this information is generated from the learners' own texts, providing these levels signifies offering them external representations that can unbias the internal representation of own and others' cognitive characteristics. The assumption is that their provision positively affects metacognitive processes. On the one hand, providing such levels could improve cognitive processing as it might prompt learners to better estimate their own cognitive characteristics and to better model the learning partners' knowledge (condition for better metacognitive processes, see section 1.1.1.). On the other hand, as grouping assures knowledge heterogeneity, providing levels could improve communication. The provision of such information might increase the chances of prompting learners to ask questions on topics with low levels of own knowledge or on topics with learning partners' higher knowledge and to explain more or better on topics with high levels of knowledge or rather higher levels than the learning partner (condition for better communication, see section 1.1.2.). In summary, the combined text mining-based grouping and representing can generate a graphic with bar charts for each pair of learners (see examples in Erkens et al., 2016 and Erkens & Bodemer, 2019), which not only involves little work for the instructor but also should effectively promote the learning of co-learners.

1.3.4. Summary and research questions

Text-mining methods have the potential to facilitate the instructors' work and improve the learners' knowledge exchange by providing information about their cognitive characteristics. Regarding instructors' facilitation, further research is necessary for analyzing and selecting text-mining methods. Instructors can be supported in guiding the learners because applying text-mining methods yields (1) distance values that can represent heterogeneity and are usable for group formation, and (2) concept clusters that are interpretable as topics (higher level concepts) and values per cluster, both of which can be represented in visualizations to improve group awareness. However, as previously described, several approaches are available; hence, text-mining methods must be selected that best meet the requirements of accurately transforming cognitive information to optimally support knowledge exchange. In addition, the selected text-mining methods must be tested in practice. These points lead to the first two research questions:

RQ1: Which text-mining methods are the most appropriate for automated grouping and representing?

RQ2: How do text mining-based grouping and representing affect learning outcomes?

Further research is also necessary with regard to the improvement of learners' knowledge exchange. The information provided in group awareness support can be classified into two, namely information about learning content (e.g., topics of learning modules) and information about learning partners (e.g., how much know learners about the topics), the influence of which on learning processes has not yet been investigated separately. The adoption of text-mining methods increases the importance of this investigation because it introduces new possibilities to generate both types of information possibly influencing cognitive processing and communication. On the one hand, text-mining methods can help to specify and provide bottom-up information about the learners' knowledge content. Such information might more effectively activate prior knowledge than top-down specified information about the content, facilitate knowledge integration, and improve the manner of addressing the content in questions and explanations. On the other hand, text-mining methods can assist with providing different types of information. The influence of both types of information on learning processes and learning outcomes requires further examination. The following questions address this research gap:

RQ3: How does the provision of information about learning partners and/or learning content influence communication?

RQ4: How does the provision of information about learning partners and/or learning content influence learning outcomes mediated by cognitive processing?

1.4. Summary of included studies

The previous section (section 1.3.) introduced the use of text mining as an efficient measure for improving the knowledge exchange between learners and relieving instructors from the effort associated with learners' guidance. The fact that people for learning have to undergo through specific cognitive and metacognitive (see section 1.1.1.) as well as communication processes (see section 1.1.2.), which usually do not simply run like this, necessitates the instructional guidance of their knowledge exchange. To guide these people, instructors can select from proven approaches such as group formation, which is a script mechanism (see section 1.2.1.), and group awareness support, which is an implicit guidance approach (see section 1.2.2.). Either way, both approaches require instructors to have cognitive information about the learners to successfully employ instructional guidance, which can be facilitated by learning analytics. Against this background, text-mining methods appear to be enriching because they facilitate the extraction of cognitive information from given texts (see section 1.3.1.). Furthermore, text-mining methods allow for transforming the collected information so that it can be a basis for instructional guidance via group formation (see section 1.3.2.) and group awareness support (see section 1.3.3.).

In this context, the studies included in this thesis examine the potential of text-mining methods in supporting knowledge exchange in collaborative learning, particularly in reducing the instructors' work when applying group formation and group awareness support and in improving the learners' cognitive processing and communication. Table 1 (p. 27) illustrates how the studies address specific research gaps related to these aspects. As discussed in sections 1.3.2. and 1.3.3, text mining is a new approach in this context, and text-mining methods need to be selected for facilitating the instructors' application of instructional guidance. For this purpose, Study 1 compared different text-mining methods in terms of their suitability for transforming cognitive information to facilitate the instructors'

Table 1. Overview of the three studies, including the current status, innovations of this work, analysis focus, and publications

Study	Current status	Innovation	Analysis focus	Paper
1	• Automated analyses of learners' interactions with learning content to group learners • Validation of the text-mining method's suitability for identifying content from learner-generated texts • Automated assessment of the cohesiveness of learner-generated content (and its visualization)	• Automated analysis of learner-generated content to group learners → **comparison of methods** • Automated analysis to identify content from learner-generated texts → **comparison of methods** • Automated assessment of the size of learner-generated content → **comparison of methods**	Automation of explicit and implicit guidance through text mining (RQ1)	**Paper 1:** Improving collaborative learning in the classroom: Text mining based grouping and representing
2	• Grouping in implicit guidance by random or by following induced characteristics • Specification of content (top-down) by instructors and representation as topics • Knowledge levels per topic are induced or based on learner input on predetermined topics • Non integration of (text mining-based) group formation on the basis of bottom-up heterogeneity and group awareness support	• Group formation based on given heterogeneity identified by text-mining methods • Specification of content (bottom-up) by text-mining methods as a basis for visualizing topics • Knowledge levels per topic are generated by text-mining methods • Design of an integrated model and text mining-based tool for group formation and group awareness support → investigation of the effects of text mining-based guidance on **learning outcomes**	Guiding knowledge exchange in (computer-supported) collaborative learning (RQ2)	**Paper 1:** Improving collaborative learning in the classroom: Text mining based grouping and representing
3	a. No separate investigation of the effects of representing information about learning partners and learning content on the length of communication	a. → Investigation of the separate and combined effects of representing (text mining-generated) information about learning partners and learning content on the length of communication **(communication)**	Mechanisms of knowledge exchange in collaborative learning (RQ3)	**Paper 2:** Which visualization guides learners best? Impact of available partner- and content-related information on collaborative learning
	b. No separate investigation of the effects of representing information about learning partners and learning content on topics addressed in communication, and learning outcomes mediated by metacognitive processing	b. → Investigation of the separate and combined effects of (text mining-generated) information about learning partners and learning content on the topics addressed **(communication)** and **learning outcomes** mediated by knowledge integration **(cognitive process)** and partner modeling **(metacognitive process)**	Mechanisms of knowledge exchange in collaborative learning (RQ3 and RQ4)	**Paper 3:** Improving collaborative learning: Guiding knowledge exchange through the provision of information about learning partners and learning contents

work of grouping and representing. In addition, the selected text-mining methods needed to be tested in practice to ascertain their capacity to improve learning in knowledge exchange. Therefore, Study 2 applied a tool that combined the selected text-mining methods for grouping and representing and investigated its overall effect on learning outcomes in the classroom. The aforementioned group awareness support intertwined the provision of two types of cognitive information, namely information about learning content and information about learning partners, the influence of which on learning processes has not yet been investigated separately. Hence, the importance of recognizing the mechanisms that could be triggered by the different types of information, particularly when the application of text-mining methods promises to offer additional benefits, was further highlighted. To gain an increased understanding of the underlying mechanisms, Study 3a investigated separate and combined effects of providing both types of information on communication with regard to the length of explanations. Study 3b also explored separate and combined effects of providing both types of information on (1) communication (topics addressed in knowledge exchange), (2) cognitive processing (verbalized knowledge integration and partner modeling), and (3) learning outcomes. For a systematic analysis in Study 3, text-mining methods were simulated in a laboratory setting. The next subsection summarizes the key contributions of all three studies.

1.4.1. Study 1: Analysis and selection of text-mining methods

To apply group formation and group awareness support based on cognitive characteristics, instructors need to collect cognitive information, which can be facilitated by automating the processing of information from text with text-mining methods. However, as argued in section 1.3., numerous text-mining methods are available for grouping, which can be based on the analysis of text closeness, and representing, which can be based on the assessment of concept closeness. These methods can build on different models, which can be roughly divided into VSM and PTM. They primarily differ in the fact that clusters resulting from VSM-based methods are distinct, whereas PTM-based methods allow concepts to appear in several clusters. Thus, a decision on the specific methods to be selected in a text mining-based tool for facilitating group formation and group awareness support is necessary. For this purpose, Study 1 investigated the question of the most appropriate text-mining methods for automated grouping and representing (RQ1). For more details, see *Text mining as a basis for forming groups and representing cognitive information* in Paper 1 (Erkens et al., 2016).

To answer the first research question, proven VSM- and PTM-based clustering methods were compared by applying the methods to a self-created text corpus with predefined content properties. Therefore, five essays about global warming were prepared, which covered 11 subtopics. These topics were distributed across the texts (ranging from “topic occurs in one text” to “topic occurs in each text”) to establish various differences between each pair of texts: topic-related differences (ranging from “no similarities” to “very high similarities”) and differences of opinion (given or not). These human-defined content properties were compared with the results of the respective automated analyses to select the methods that most accurately capture the manually defined properties of the text corpus. For more details, see *Design of the Grouping and Representing Tool (GRT)* in Paper 1 (Erkens et al., 2016).

The results indicated that the PTM-based concept clustering with LDA was more suitable for detecting topics from a given text corpus (as a basis for specifying the content of knowledge) than the VSM-based concept clustering, as it identified all 11 topics (in contrast to the VSM-based clustering that identified only eight topics). For generating the extents of topics in each text (as a basis for representing the knowledge levels), the PTM-based LDA also performed better in the systematic comparison and reproduced the ratio of the topics’ extents closely to human estimate. Regarding text clustering to detect the distance values between texts (or rather learners), the VSM extended by Euclidian distance outperformed the distance metrics resulting from LDA because it was more similar to the text differences (as a basis for group formation) determined by humans. Based on these results, two decisions were made for a text mining-based *Grouping and Representing Tool*. (1) Concept clustering, which forms the basis for the representation of information about learning content in the form of topic lists and of information about learning partners in the form of bar charts, is based on LDA. (2) Text clustering, in the course of which a vector distance metric is necessary to describe the distance between vectors, is implemented with a VSM combined with the generation of Euclidean distance values. This metric also forms the basis for the formation of learning groups, starting with the grouping of learners with the highest values and then downwards. Figure 5 (p. 30) presents an overview of the methods selected for text mining-based grouping and representing from Study 1. For a more detailed description of the text mining-based *Grouping and Representing Tool* with its different functions, see also *Specification of the functions of the GRT* in Paper 1 (Erkens et al., 2016).

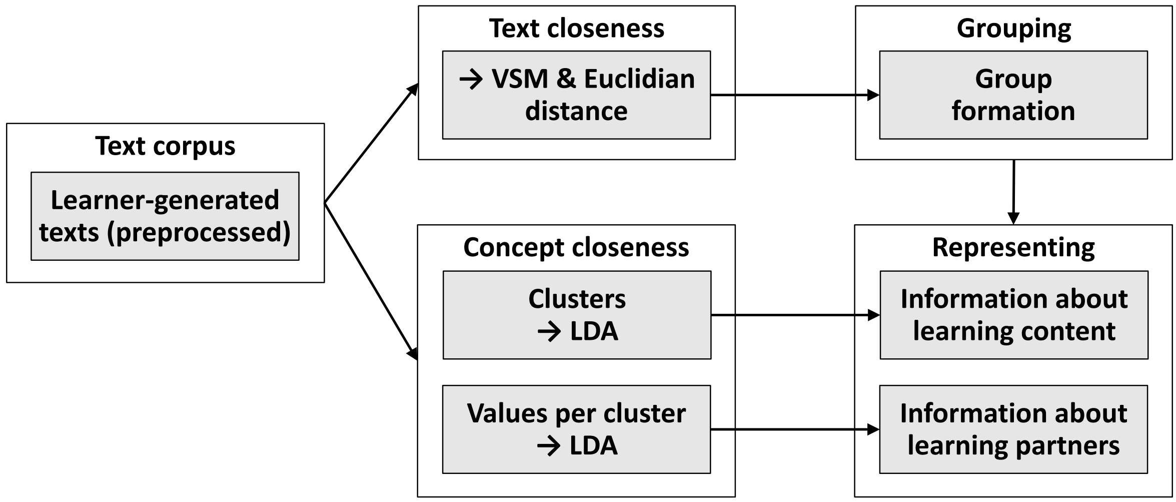

Figure 5. Information processing in the *Grouping and Representing Tool*. The input, preprocessed texts (left box), is the basis for (1) calculating Euclidian distances (top middle box) to inform group formation (top right box) and (2) clustering concepts based on LDA (bottom middle box) to facilitate the representation of information (bottom right box).

1.4.2. Study 2: Effect of selected text-mining methods in a combined guidance tool

In addition to establishing their suitability for facilitating the instructors' work, the selected text-mining methods had to be tested for their appropriateness for improving collaborative learning. As argued in section 1.2, instructional guidance approaches such as group formation and group awareness support can individually promote learning in knowledge exchange. In the context of developing a text mining-based support for both approaches, combining them to create added value was reasonable. Based on the selection of methods for the resulting *Grouping and Representing Tool* in Study 1, a need to examine the influence of this combined guidance support in practice emerged. Therefore, Study 2 investigated in a school scenario how text mining-based grouping and representing affects learning outcomes (RQ2). For a detailed description of the hypotheses, see the *Dependent variables & hypotheses* section in Paper 1 (Erkens et al., 2016).

To answer the second research question, Study 2 applied the text mining-based guidance in a real classroom setting. To validate the overall effect of the *Grouping and Representing Tool* on learning, 54 high school students of a German upper secondary school with a main focus on geography were examined in a 2×2 mixed factorial design. The between-subject factor was the randomly assigned experimental condition (tool support vs. no tool support), whereas the within-subject factor was the respective phase in the procedure of collaborating (phase of writing an essay vs. phase of modifying the essay). In the writing phase, learners were instructed to write an individual essay about the topic “Global warming: what is the extent of its natural and man-made causes? What countermeasures can be taken?” The students were subsequently assigned to dyads to exchange their knowledge with their learning partner, either supported by the tool or not. In the modifying phase, the students should individually modify their essays, if the exchange of knowledge with the learning partner resulted in new knowledge. Modifying was more precisely defined as deleting, rewriting, or completing passages of the essay. Learning outcomes were examined by the dependent variables *learning* (sum of the topics’ extents in the essay per phase) and *knowledge convergence* (heterogeneity between co-learners per phase) with the learning partner. For more details, see *Experimental study* in Paper 1 (Erkens et al., 2016)

The influence of the tool on learning outcomes (RQ2) was investigated by two-factorial repeated-measures ANOVAs. Furthermore, a moderation analysis was performed to test whether the heterogeneity between learning partners has a stronger effect in the experimental group with tool support. Therefore, the assigned experimental condition (tool support vs. no tool support) was used as an independent variable, *heterogeneity* (text dissimilarity) as a moderator, and *knowledge acquisition* (topics’ extent from the writing phase subtracted from topics’ extent in the modification phase) as a dependent variable. To obtain further insights into the text modifications that were assumed as an indicator of learning processes, these analyses were complemented by an exploratory content analysis of the essays of the dyad with the highest knowledge convergence. Figure 6 (p. 32) illustrates a schematic representation of the object of investigation to answer RQ2.

The results confirmed the effect of the text mining-based *Grouping and Representing Tool* on learning outcomes for the most part. Students supported by the tool, which had grouped them with topic-distant classmates and provided them with information about the learning partner and learning content, added twice as many concepts in their essays (30% increase) as those students without support (15% increase). Furthermore, the dyads of learners who were supported by the tool demonstrated a higher knowledge convergence between writing and modifying the essay (18% decrease of heterogeneity) than dyads without tool support (6% decrease of heterogeneity). Although a positive relationship seemed to exist between heterogeneity and learning outcomes, the effect of the tool on knowledge acquisition did not become larger the greater the heterogeneity within dyads; hence, group formation did not achieve the desired effect. A treatment check confirmed this

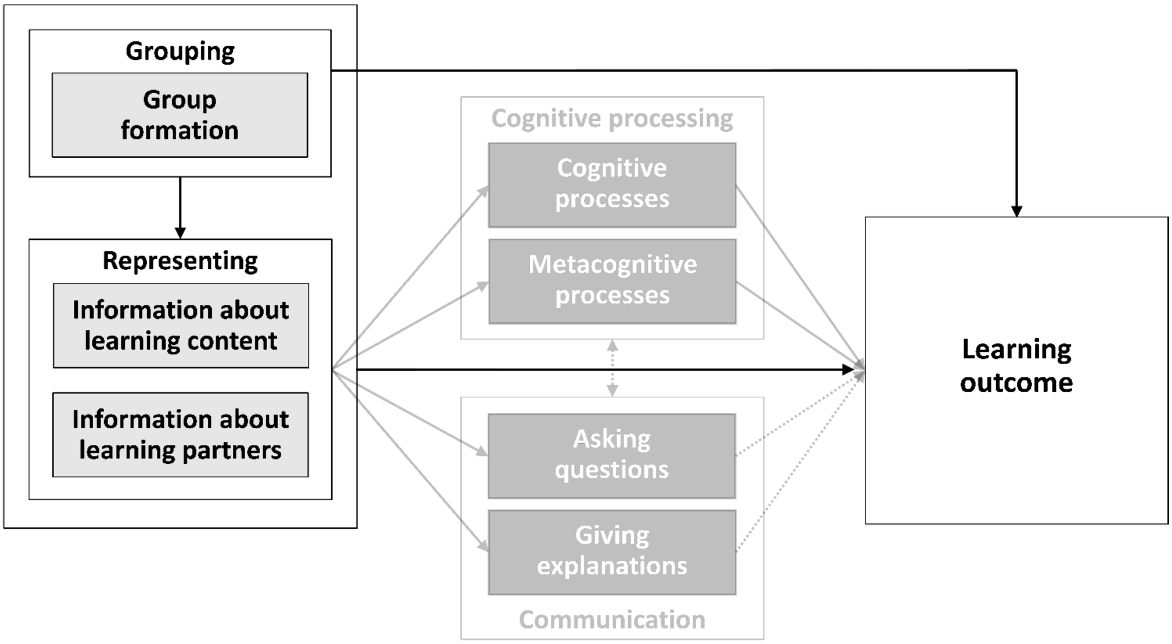

Figure 6. Focus of analysis to answer RQ2. This research question refers to the overall effect of text mining-based grouping and representing on learning outcomes as well as the effect of heterogeneity in particular.

assumption, which revealed that the distance values of dyads in the experimental group were not higher than the distance values of dyads in the control group (for more details, see *Results* in Paper 1 in Erkens et al., 2016). Based on these findings, the effect of the text mining-based *Grouping and Representing Tool* seemed to be mainly due to the different types of information provided (topics as information about learning content and bar charts as information about learning partners in a group), which might implicitly guide the learners' cognitive processing and communication.

In addition, the supplementary exploration of essay modifications of the tool-supported dyad with the highest knowledge convergence identified two types of adaptations, which might explain the aforementioned effects. These two types include (1) the adaptation of one's own text to topics of the whole group, and (2) the adaptation of one's own text to the learning partner's text. In case (1), the adaptation involved the addition of concepts that were provided or related to provided topics and were largely inconsistent with the concepts of the partner text. This type of adaptation might refer to cognitive processing; the representation of topics might have caused the activation of concepts from prior knowledge, which were subsequently integrated into the own essay. In case (2), the adaptation entailed the addition of concepts related to the list of topics but also included in the partner text in a similar form. A further characteristic for this type of adaptation was that the added concepts were integrated in the text effectively, as they were linked to concepts from other topics. This type of adaptation might pertain to the selection of topics and cognitive elaboration. As concepts were added in cases indicating that both the learner and the partner know nothing or the partner knows more, the learner reasonably focused on own knowledge gaps in questions, regardless of the other's knowledge level. As the learner further often linked the added concepts to other topics in the own text, these concepts are assumed to have been cognitively elaborated in the exchange, thus simplifying their integration. Across cases, the exploratory observation eventually demonstrated that learners preferred to exchange about listed topics on opinions or everyday knowledge rather than about topics on specialist knowledge, which might be a disadvantage of the text mining-based, bottom-up approach that might allow for an over-representation of everyday knowledge.

In summary, as the effectiveness of group formation was challenged, the effects of the *Grouping and Representing Tool* on learning seemed to be mainly due to providing the two different types of text mining-generated information, topics (information about learning content), and the extent of topics in the text (information about learning partners). This conclusion confirmed the intention to further investigate the confound of both types of information in Study 3.

1.4.3. Study 3: Mechanisms underlying (text mining-based) guidance

In addition to verifying the general effect of text mining-based guidance as in Study 2, increasing the understanding of the mechanisms underlying guided knowledge exchange is important. As argued in section 1.1., specific modes of cognitive processing and communication are particularly beneficial for learning in the exchange of knowledge. Understanding these modes might also enhance the use of text-mining methods, the application of which could even introduce new means of positively influencing prior knowledge activation (see section 1.3.2.). Based on this rationale and the findings of Study 2 that (1) the grouping algorithm must be redesigned and (2) the types of information seem to have different effects on cognitive processing and communication, the focus was directed to implicit guidance through representing. To increase the understanding of the guidance mechanisms through awareness support, disentangling the effects of different information types in the presented visualizations was necessary. In Study 2, as in previous research in this area (see section 1.2.2.), information about learning partners (available bar charts representing levels of prior knowledge in Study 2) and about learning contents (specification of content of prior knowledge in a topic list in Study 2) were not examined separately with regard to their influence on cognitive processing and communication. Thus, a general research gap exists in terms of the separate and combined effects of representing both information types, which is also particularly important for the design of the text mining-based support, as an added value can be assumed through a possibly better activation of prior knowledge. To close this research gap, Study 3 investigated the influence of the provision of information about learning partners and learning contents on learning processes (RQ3 and RQ4) and learning outcomes (RQ4), using the simulation of a text mining-based representation of both types of information. For a detailed description of the hypotheses, see the section on *Instructional purposes of cognitive group awareness tools* in Paper 2 (Erkens & Bodemer, 2017) and *1.4. Hypotheses* in Paper 3 (Erkens & Bodemer, 2019).

To answer research questions 3 and 4, Study 3 separated the provided information about learning partners and information about learning content in an experimental laboratory setting. To investigate the separate and combined effects of the respectively provided information on learning processes and learning outcomes, 120 university students were randomly assigned into experimental groups of a 2×2 between-subjects factorial design. In this arrangement, the availability of information about the learning partner (available vs. unavailable) and the specificity of information about the learning content (specified vs. unspecified) was varied. Dependent on the respective experimental groups, the participants were provided with one out of these four versions of visualizations that illustrated their prior knowledge on climate change and bioenergy:

(1) Visualizations with self-related, no partner-related, and no content-specific information;
(2) Visualizations with self-related, partner-related, and no content-specific information;
(3) Visualizations with self-related, no partner-related, and content-specific information; and
(4) Visualizations with self-related, partner-related, and content-specific information.

The respective versions of visualization were provided to the participants during two phases of a bogus collaboration, and they had some specific characteristics dependent on the phase (see all the versions of visualizations in section *2.3. Procedure and instructions* in Paper 3, Erkens & Bodemer, 2019). In the first phase of collaboration, the visualization was adapted to a simulated scenario in which the participants were induced to have a lower total level of prior knowledge than their learning partners. In this phase, the participants were instructed to write questions to their first learning partner. In the second phase of collaboration, the visualization was adapted to a simulated scenario with another learning partner in which the participants were induced to have a higher total level of prior knowledge than their learning partners. This time, they were asked to write answers to three questions of their second learning partner: What are (1) the advantages of bioenergy, (2) the disadvantages of bioenergy, and (3) their conclusions on bioenergy in the context of climate change? The prior knowledge of the participants that was presented in the visualization was induced by a text on the advantages of bioenergy in the first collaboration phase. This text could be accessed during both collaboration phases. In addition, a text by the first bogus learning partner on the disadvantages of bioenergy was provided between the first and the second collaboration phases to induce the increased knowledge in the second collaboration phase. As dependent variables, communication, cognitive processing, and learning outcomes were investigated. Considering communication, the following variables were examined: *length of explanations* (Study 3a) and *number of task-relevant concepts* (topic selection) in questions and explanations (Study 3b), which was determined by content analysis. With regard to cognitive processing, learners' *partner modeling accuracy* (as a result of metacognitive processes) and the *level of knowledge integration* in explanations captured by content analysis (referencing cognitive processes) were investigated; learning outcomes were operationalized as *recall score* that determined the quality of recalling the learning partners' contribution by content analysis (Study 3b). Further details regarding Study 3a are found in the *Methods* section in Paper 2 (Erkens & Bodemer, 2017), and all the details regarding Study 3b in section *2. Material and methods* of Paper 3 (Erkens & Bodemer, 2019).

The question of how providing information about learning partners and/or learning content influences communication (RQ3) was investigated by two-factorial ANOVAs, which were complemented by an exploratory qualitative analysis in Study 3a contrasting the learners with the highest and lowest learning outcomes from the experimental groups that were provided with visualizations marked above as (1) or (4). Furthermore, Study 3b conducted two additional mixed ANOVAs, which included knowledge level as a within-subject variable to investigate its effect on *topic selection*. One analysis investigated the visualized *levels of learners' prior knowledge* (much vs. little) as a within-subject variable in addition to the between-subject variable *specificity of information about the learning content* (specified vs. unspecified). The second analysis examined the knowledge distribution, visualized *levels of learning partners' prior knowledge compared with learners' prior knowledge* (learning partner has the same prior knowledge as learner vs. learning partner has less prior knowledge than learner), as a within-subject variable in addition to the between-subject variable *availability of information about the learning* *partne*r (available vs. unavailable). Figure 7 illustrates a schematic representation of the object of investigation to answer RQ3.

Regarding questions, the results have not confirmed the expectation that providing specified information about learning content in general guides the selection of topics during knowledge exchange; however, the results indicated that such provision implicitly guides asking questions when it is combined with the provision of self-related information (*levels of learners' prior knowledge*). The expectation that learners focus own knowledge gaps (topics with little *levels of learners' prior knowledge*) in their questions (Study 3b, see section *3. Results* in Paper 3, Erkens & Bodemer, 2019) was contradicted by the finding that learners instead asked significantly more questions on topics where a high level of prior knowledge was represented. However, the results denoted that such behavior that is unintended from a guidance viewpoint was after all reduced when specified information about learning content was additionally provided. Thus, the success of addressing own knowledge gaps in questions presumably depended on the specificity of information about learning contents.

Regarding explanations, the results revealed that providing information about learning partners guides learners to give longer but not better explanations, in which "better" denotes the use of more task-relevant concepts from the visualization. Furthermore, the results

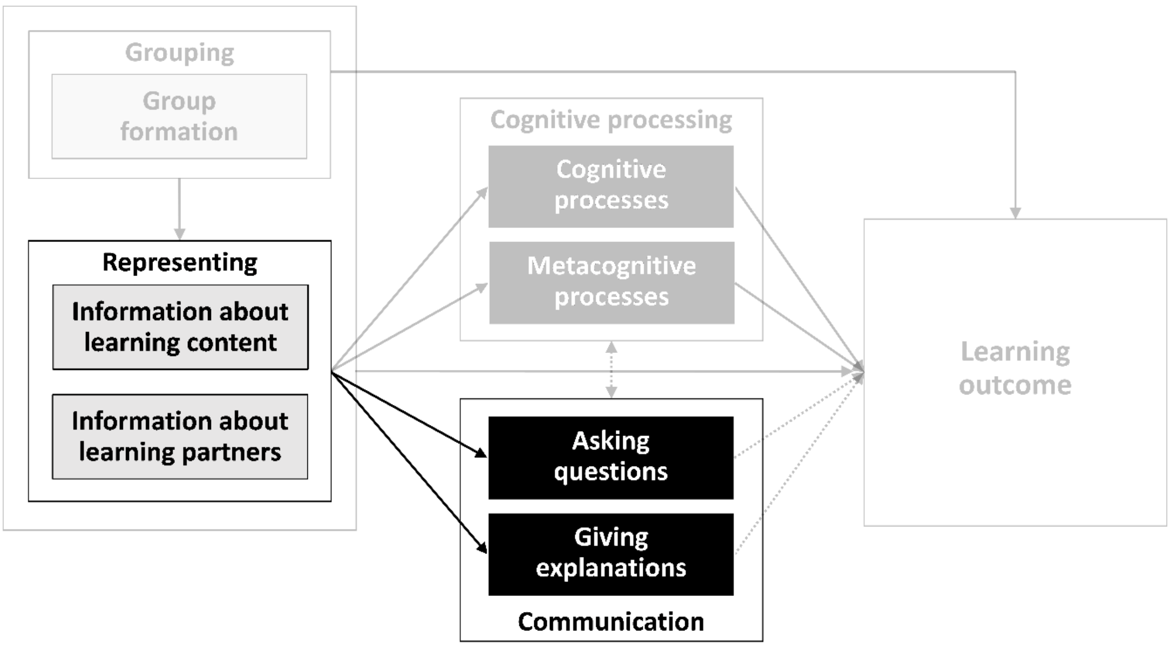

Figure 7. Focus of analysis to answer RQ3. This research question refers to the effect of provided types of information on communication with regard to the learners' topic selection in questions and explanations as well as the length of explanations.

indicated that an improved topic selection in explanations additionally depends on the available information that learning partners know less than the learners themselves. In essence, the expectation is that the effect of providing information about learning partners on explaining to others is supported when information about the learning content is also specified. Nevertheless, the results implied neither the expected interaction effect on the length of explanations (Study 3a, see the *Impact of partner-related and content-related information on explanations* in Paper 2, Erkens & Bodemer, 2017) nor on the selection of task-relevant concepts (Study 3b, see section *3. Results* in Paper 3, Erkens & Bodemer, 2019). The additional qualitative comparison that contrasted learners with the highest and lowest knowledge gains (Study 3a, see *Contrasting cases: Comparison of successful and unsuccessful learners* in Paper 2, Erkens & Bodemer, 2017) indicated that providing specified information about the learning content could still be relevant for implicit guidance because successful learners seem to use this type of information for addressing topics in their explanations. This observation also revealed another aspect: successful learners across experimental groups frequently addressed topic areas of own missing knowledge or misconceptions, but only the successful learner provided with information about learning partner and learning content also wrote frequently about specific concepts of own missing knowledge or misconceptions. Moreover, although this learner mentioned that he/she did not use the visualization to decide what to explain, the analysis revealed that this learner especially explained topics displaying that she/he is more knowledgeable than her/his learning partner. Thus, available information about the learning partner seems to trigger longer explanations but not the use of more task-related concepts; the provided information about the knowledge distribution of one's own knowledge compared to learning partner's knowledge might additionally guide the selection of topics to be explained.

Taking into account the knowledge distribution in addition to the availability of information about the learning partner in a mixed design (Study 3b, see section *3. Results* in Paper 3, Erkens & Bodemer, 2019), the results yielded the significant main effect of the visualized levels of prior knowledge. Learners mention more task-relevant concepts on topics where a learning partner's level of prior knowledge is lower than the learner's own level of prior knowledge than on topics where both have the same levels of prior knowledge. Additionally, the availability of information about the learning partner reinforced this communication behavior. In further considering the aforementioned results from Study 3a, it can be concluded that task-related explanations can especially be expected if the comparison with the learning partner or rather the ratio of the displayed bar charts demonstrates that the learning partner knows less than the learner.

A multi-categorical mediation analysis in Study 3b investigated the question of how providing information about learning partners and/or learning content influences cognitive processing and learning outcomes (RQ4). Therefore, the experimental group with only self-related information available was used as a reference group that was compared to the three other experimental groups. These three other groups are as follows: (1) the experimental group in which the learners were provided with the visualization additionally representing information about the learning partner but no specified information about the learning content; (2) the experimental group in which the learners were provided with the visualization representing no information about the learning partner but additional specified information about the learning content; and (3) the experimental group in which the learners were provided with the visualization additionally representing information about the learning partner and specified information about the content. The cognitive processing variables *partner modeling accuracy* and *level of knowledge integration* were examined as parallel operating mediators. The *recall score* determining the quality of recalling the knowledge of the first learning partner who provided the information on the disadvantages of bioenergy (see the procedure description above) served as a dependent variable illustrating the learning outcome of learners. Figure 8 (p. 39) illustrates the schematic representation of the object of investigation to answer RQ4.

The results of the mediation analysis indicated that providing information about learning partners but not information about learning contents implicitly guides cognitive processing and therewith induces learning. Only the switch from the reference group to the experimental group with available information about the learning partner but no specified information about the content indirectly improved the *recall scores* through the *level of knowledge integration* and *partner modeling accuracy* (see *3. Results* in Paper 3, Erkens & Bodemer, 2019, for an overview of all values). This outcome supports the claim that an improved *level of knowledge integration* and *partner modeling accuracy* might help learners to exchange their knowledge more successfully when the information about a less knowledgeable learner is provided without content-specification. The subsequent discussion explains the extent to which this result contradicts previous assumptions and findings of this work and how the results from this chapter as a whole are to be interpreted against the theoretical background.

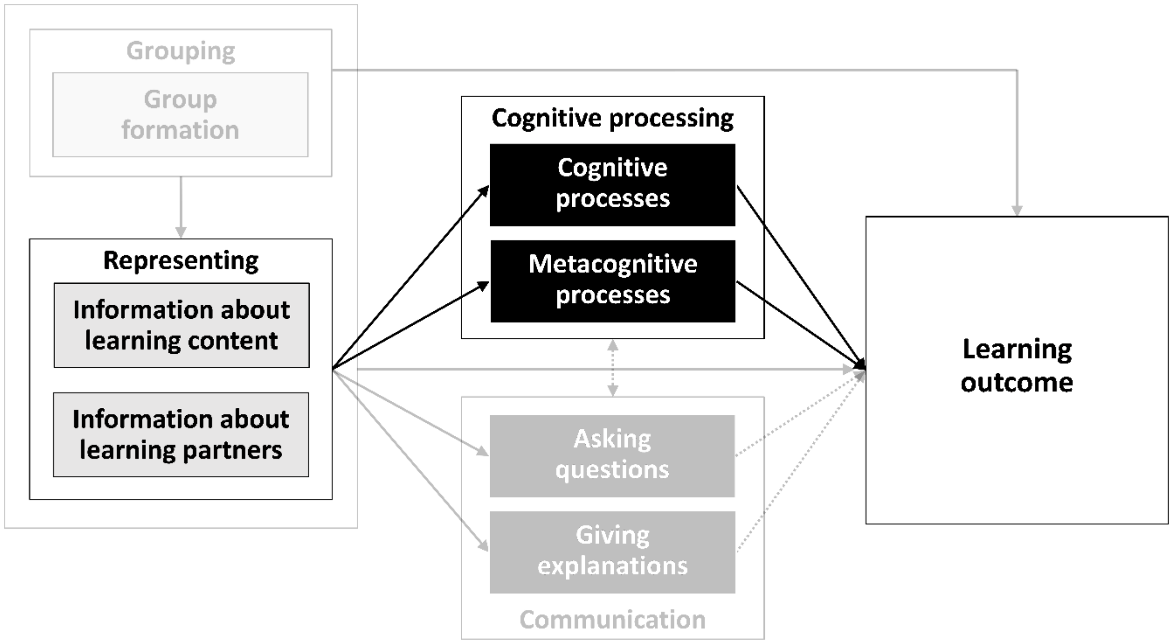

Figure 8. Focus of analysis to answer RQ4. This research question refers to the effect of the provided types of information on learning outcomes mediated by cognitive processing through cognitive elaboration and partner modeling.

1.5. Synthesis and conclusion

Text mining offers a promising potential to assist with guiding knowledge exchange in collaborative learning. Study 1 highlights the appropriateness of specific text-mining methods for facilitating the instructors' work of grouping and representing (RQ1). Combining VSM with Euclidian distance is the most accurate method to inform group formation, whereas LDA suitable is for facilitating group awareness support. Using the selected text-mining methods from Study 1 and integrating their functions of grouping and representing, Study 2 shows their effectiveness with regard to instructional guidance (RQ2). Text mining-based group formation and group awareness support guide learning partners to learn more than learning partners without this support. To complement the insight from Study 2 that especially providing information about learning partners and about learning content has a guiding effect, Study 3 reveals how offering these different types of information influences communication (RQ3): providing self-related information increasingly triggers questions on topics with knowledge gaps when the content is specified. Available information about the learning partner triggers the adaptation of explanations, such as contributing more and addressing the knowledge gaps of the learning partner. Beyond its influence on communication, Study 3 further reveals how offering different types of information influences cognitive processing and learning outcomes (RQ4): even though the provision of specified information about the content has not the expected effect, the availability of unspecified information about learning partners seems to trigger the cognitive process of elaboration, resulting in knowledge integration in explanations, and the metacognitive process of more accurate partner modeling, both improving learning. Figure 9 presents an overview of the results of all the papers included, which are discussed in the next section in an integrated manner that transcends the individual discussion of the papers.

1.5.1. Discussion of results and practical implications

The studies conducted highlight not only the potential of text mining in supporting knowledge exchange in collaborative learning but also the necessary adjustments prior to its practical use. The comparison of text-mining methods (Study 1) mostly confirms that specific methods are suitable for collecting information for informing group formation and facilitating group awareness support. With regard to informing group formation, VSM-based clustering with Euclidean distance can accurately reproduce the ranking of the knowledge differences of prepared essays, while the LDA is less accurate. An explanation for this result cannot directly be drawn from preliminary work because method comparisons are sparse in

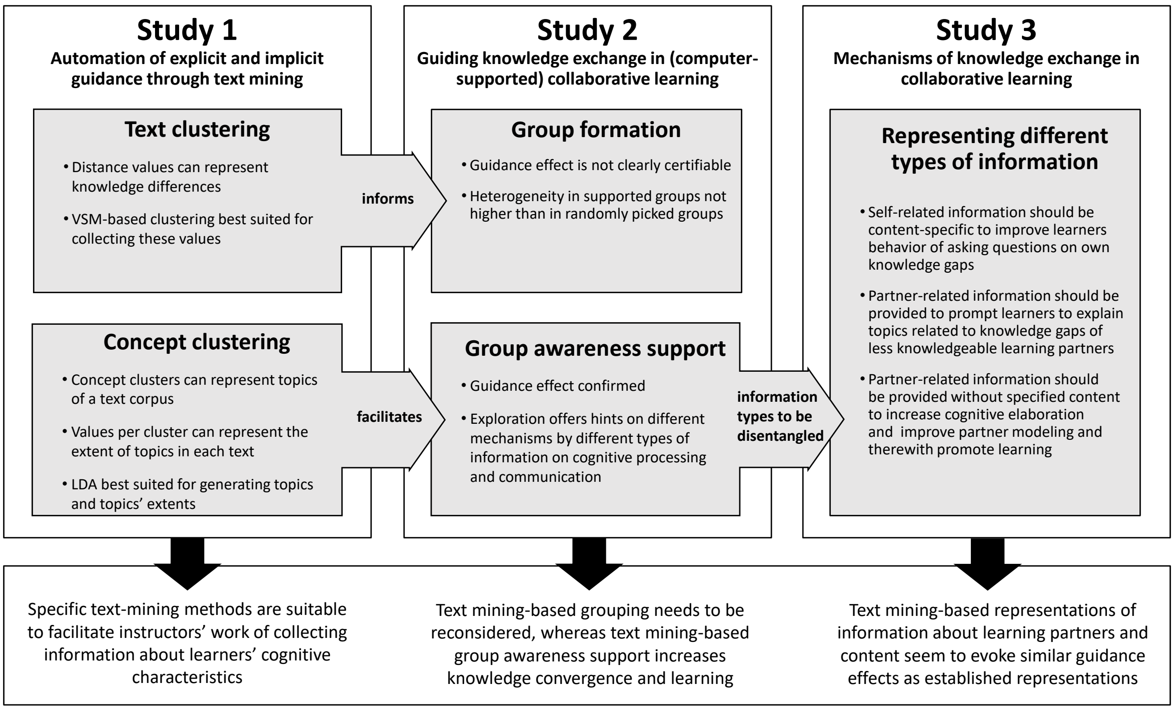

Figure 9. Study results and practical implications. Study 1 results in the selection of text-mining methods informing the functions of the *Grouping and Representing Tool*, Study 2 offers hints on the tools' optimization potential and the single effects of the information types provided, and Study 3 presents new insights into the mechanisms of (text mining-based) guidance.

this context. However, from a methodological viewpoint, VSM-based clustering might be the superior method due to its high level of preciseness. This method considers each text as a vector, including concepts from the entire text corpus for the analysis, whereas the adapted LDA seems to promote inaccuracies without such one-to-one matching. In addition, neither one method nor the other can identify different opinions expressed in the texts, which implies that the heterogeneity determined here can support learners in the Vygotskian (Vygotsky, 1978) rather than the Piagetian (Piaget, 1977) sense.

Regarding the facilitation of group awareness support, text-mining methods can identify the content of systematically produced essays to the most part (in the case of using a VSM-based factor analysis) or even to their entirety (in the case of using LDA). This point is consistent with Sherin's (2012) result that the VSM-based clustering of concepts is suitable for identifying various topics from texts. However, the present study indicates that the completeness of the topic identification is important (see also the limitations in section 1.5.2), which Sherin (2012) did not consider in his exploratory work. Vector space model-based clustering does not achieve such completeness, but LDA does, which has been successfully used for identifying topics (Southavilay et al., 2013). Furthermore, LDA also proves to be a better method for capturing the extent to which topics occur in each text, which is not surprising, as it also identifies the topics more effectively. Thus, keeping the aforementioned issue of completeness in mind, the preliminary practical implications from this methodical pre-study involve (1) informing group formation by the results of VSM-based clustering with Euclidian distance, and (2) applying group awareness support based on LDA. Both implications can be easily combined to enhance individual effects, but they must first pass a field test to be applied.

In practical testing (Study 2), text mining-based guidance seems to help people to exchange their knowledge more effectively and learn further. With text mining-based guidance, learners add more new aspects to their essays after a discussion with a learning partner than without this support. This result fits in with the numerous findings demonstrating the guiding effect of knowledge-complementary group formation (see Johnson et al., 2000) and group awareness support (see Bodemer et al., 2018), and it suggests that a similar effect can be achieved with text mining-based guidance. Additionally, the high knowledge convergence of supported learners indicates that text mining-based guidance might also meet the goal of developing a shared understanding, which is another objective of collaborative learning (e.g., Dillenbourg, 1999; Roschelle & Teasley, 1995; Scardamalia & Bereiter, 1994). However, knowledge convergence and better learning outcomes cannot be clearly attributed to both guidance approaches, although the integration of these approaches into one tool is a new aspect here to potentiate the effects on learning. Attributing the effects to group formation is not possible because of the lack of higher knowledge heterogeneity within learning groups supported by the *Grouping and Representing Tool* in comparison to unsupported learning groups (see also the limitations in section 1.5.2). On the one hand, this inadequate level of heterogeneity in the supported learning groups may be due to the text-mining method used; problems with the methods may not have been observed in Study 1 because of other prevailing conditions. On the other hand, the inadequate level of heterogeneity can be due to the small size of classes; for instance, a group of less than 30 members may pose more difficulty in identifying high knowledge differences between students than a group of 300 individuals, especially if they all attend the same classes and their knowledge might be very similar. Either way, informing about text mining-gathered topics and extents to which these topics occur in texts seems to have a strong influence on knowledge exchange and learning. Consequently, text-mining methods for group formation need to be reconsidered and further examined before they can be used in practice to reinforce the effects, whereas text mining-based group awareness support seems appropriate for use.

Equally important, the exploratory analysis of the texts from the learning partners with the highest knowledge convergence indicates that text mining-based support of group awareness might promote prior knowledge activation and/or the selection of concepts. Adapting the text to topics of the entire group or rather adding topic-related concepts to an essay fits in with the assumption that providing higher level concepts activates prior knowledge (Ausubel, 2000). As the concepts added to the own essays were related to the listed topics but mostly not consistent with the concepts explained in the learning partner's essay, the assumption is supported that text modifications are due to the activation of prior knowledge and not due to the exchange with the learning partner. This observation might also support the conclusion that the activation is promoted by text mining's bottom-up approach of identifying topics that the whole group of learners associates with the learning task. However, without knowing the discussion between the two learners, the point that adding the concepts in the text might be due to the exchange of knowledge with the learning partner cannot be excluded completely. In this case, the provided information could at least have guided the selection of the topics of the exchange. However, the attribution of this selection to the listing of the related topics and concepts or to the visualization of related levels of knowledge remains unclear.

Next, the exploratory analysis indicates that text mining-based support of group awareness might promote the selection of topics and cognitive elaboration during the exchange. Adaptations to the learning partner's essay or rather adding concepts from the partner's text to the own text implies that questions have been asked on these topics during the exchange. This explanation would suit previous findings from group awareness research that new knowledge is acquired in cases of knowledge gaps regardless of the knowledge level that was displayed for the learning partner (Dehler et al., 2009). Moreover, linking the added concepts to other topics in the own text suggests that the concepts have been cognitively elaborated in knowledge exchange. In previous research, such cognitive elaboration in knowledge exchange could be traced back to informing learning partners about higher knowledge levels of one learner (Dehler Zufferey et al., 2011); learners who are aware of their differences primarily exchange about unshared knowledge (Schittekatte & Hiel, 1996) and thereby might more strongly converge their knowledge, which is also given in this example case. However, whether this elaboration can be traced back to a higher level of knowledge on the part of the learning partner is uncertain, as a higher partner level was not the case with all the added concepts, only with some. Furthermore, conclusions about knowledge exchange in this exploratory analysis can only be made on the basis of text modification and only for one dyad; thus, an investigation of the guiding effects of text mining-based representing on communication is pending, which is part of Study 3.

With regard to the guidance of communication (Studies 3a and 3b), different mechanisms can be triggered by providing (text mining-generated) information about the learning partners and/or information about the learning content. Considering questions, learners seem to ask their learning partners more questions about topics with own knowledge gaps if the available self-related information is content-specific. This inference not only supports the above-mentioned exploratory observations from Study 2 but it is also consistent with earlier research where both types of information were provided in combination, and in this combination raised questions about content with knowledge gaps (Dehler et al., 2009). Another argument is that the provision of the content-specific information about one's own knowledge gaps activates metacognitive skills (see Efklides, 2008), hence promoting learning. Without specified information on the content, however, learners primarily ask questions on topics where a higher level of knowledge is indicated, which does not fit with previous findings and is even contrary to expectation. An explanation for this postulation might be the missing reference to the learners' actual knowledge gaps. In support of the systematic variation in this study, the provided content simulated rather than represented the individual prior knowledge of learners, and this missing authenticity of provided information might have biased the results (see also the limitations in section 1.5.2.). In summary, visualizations based on the text-mining simulation seem to guide questions similarly to other established visualizations. Overall, self-related information should be content-specific to guide questions.

Regarding answers to questions, learners seem to provide longer but not thematically adapted explanations if (text mining-generated) information about the learning partner is available. The expected added value of providing specified information on learning content, which should be revealed by disentangling both types of information, is not supported. The length of explanations and the use of concepts from the visualization seem to be independent of such specification. Such communication behavior would have been an indicator for an increased activation of prior knowledge through the visualization of higher level concepts (see Ausubel, 2000). As this is not the case, an advantage of the text mining-based bottom-up approach, which was already scrutinized in Study 2 for possibly directing the learner's focus on everyday knowledge (see also the limitations in section 1.5.2), cannot be confirmed. The adaptation of explanations depends on the learners being informed that the learning partner knows less, as learners in the case of such knowledge distribution are more likely to address topics with the learning partners' knowledge gaps in their explanations. This result is consistent with previous findings that also describe such an adaptation in the sense of audience design (see Clark & Murphy, 1982; Lockridge & Brennan, 2002; Schober & Brennan, 2003), when information about a less knowledgeable learning partner is available (Dehler et al., 2011; Dehler et al., 2009). This premise is also supported by the exploratory analysis in Study 3a: the most successful learner displays exactly the adapted behavior of topic selection in explanations as described above. However, this learner was also supported with specified information about the content in addition to visualized knowledge distributions that he/she seemed to use as a reminder of the concepts that are relevant for understanding a topic. However, a general effect of text mining-based bottom-up approach cannot be confirmed. Hence, visualizations based on text mining-simulated information about the learning partner seem to guide explanations similarly to other established visualizations. In all the cases, the possible comparison of one's own and the learning partners' levels of knowledge seems to be important.

In terms of the (text mining-based) guidance of cognitive processing (Study 3b), offering plain information about the learning partner (only two bars for comparison on the main topic) seems surprisingly to be the best means of improving learning processes and learning outcome. Providing non-specific information about learning partners enhances cognitive elaboration and partner modeling accuracy and promotes learning. This effect is not in line with the assumption based on the exploratory analysis from Study 2 that specified information about the content might be of relevance for knowledge activation and thus for integrating concepts (whereby it should be noted that in Studies 2 and 3, conclusions on cognitive processes were only drawn on the basis of text contributions, see the limitations in section 1.5.2.). Moreover, the results do not exactly fit with the assumptions based on previous studies, where similar effects were found with the availability of both types of information (Dehler Zufferey et al., 2011; Sangin et al., 2011). A new insight in the current research is that the mentioned positive effects occur especially if the information about learning partners is not content-specific, which is contrary to the expectation that providing specified content has an added value. This contradictory finding may be due to a larger number of provided concepts than in the aforementioned studies, possibly overloading people's cognitive systems (see Merriënboer & Sweller, 2005). Such overload could hamper cognitive elaboration, which is indicated by the cognitively demanding integration of knowledge in explanations, and cause the inaccurate assessments of partner knowledge, which should be avoided by specifying the content (see limitations in section 1.5.2.).

By contrast, the success of providing plain information about the learning partner might be grounded in several facts. Regarding cognitive elaboration, the adaptation of explanations to a recipient also depends on the motivation to self-evaluate through social comparison (Ray et al., 2013); knowledgeable explainers who are motivated to engage in social comparison, which is facilitated by group awareness support, might prefer to keep their distance from the learning partner by withholding information (Ray et al., 2013). Thus, plain bars that are based on aggregated values and exhibit higher knowledge differences might not present learners with the danger that their learning partner could catch up excessively and "permit" better elaboration. Another explanation would be that the plain information on a less knowledgeable learning partner increases learners' awareness of own strength and the learning partner's weakness, and thereby prompts them to assume the role of an expert. Such role assumption is called "emerging role" in collaborative learning (Kollar et al., 2006) and characterized by guiding the learners' individual behavior (Hare, 1994), which could also explain the increase in cognitively elaborated explanations. Both explanations nevertheless underline once more the importance of complementary distribution in knowledge heterogeneity to support reciprocity. With regard to partner modeling, the extent to which learners have used the provided visualization or biased strategies is unclear, such as transferring the knowledge distribution on the main topic to each subtopic or modeling the learning partners' knowledge according to one's own (see Nickerson, 1999). In the plain information condition, such biased strategies may have been particularly well supported and yielded more accurate results.

In summary, the present work highlights text mining's promising potential to support knowledge exchange in collaborative learning. Although VSM-based group formation should be reconsidered and adapted because the combination of group formation and group awareness support should even operate better when the maximal complementarity of learning groups is given, group awareness support can be based on LDA. The new insights into guidance mechanisms can be used for adjusting this support: (1) to promote questions on knowledge gaps, self-related information provided should be content-specific. For this purpose, a higher number of topics (concept clusters) should be identified by text-mining methods when applied for this purpose. (2) To support learners in focusing and cognitively elaborating on the knowledge gaps of their learning partners, information about learning partners should be provided without excessive content specification. This support implies to reduce the number of listed topics to facilitate the comparison of learning partners and to underscore the high differences in knowledge. Apart from the fact that these contradictory results on content specificity still need to be resolved, the text mining-based representing function can be employed. Visualizations that are generated on the basis of text-mining analyses seem to evoke similar guidance effects as other established visualizations. Even as they maintain the same guidance effects, they further have the advantage of collecting information more efficiently. This aspect also signifies their importance as an extension of existing tools, for which concrete application examples are presented in Paper 3 (see *4. Discussion* in Erkens & Bodemer, 2019).

1.5.2. Limitations

The preceding studies have some limitations. Studies 1 and 2 illustrated that text mining-based grouping and representing require adjustment. In terms of group formation, the text mining-based grouping in the experimental group in Study 2 did not result in higher distance values of learning groups than in the control group, suggesting the lack of differences between random and algorithm-based formations with regard to the knowledge heterogeneity of learning partners. Therefore, a different approach to grouping in the tool design is necessary that can identify pairs with large knowledge differences, even in school classes with 30 people or less. In addition, the assumption that high knowledge heterogeneity is constantly associated with the complementarity distribution of knowledge lacks certainty; one-sided knowledge distributions would also be possible in the case of high distance values between learning partners, which could hinder learning (see Deiglmayr & Schalk, 2015; Ray et al., 2013). For the same reason, reconsidering and adapting the method for text mining-based grouping would be appropriate.

With regard to text mining-based representing, the question of the optimal specificity of information about the learning content requires clarification in terms of the number and accuracy of topics. In Study 1, the precise number of topics in the text corpus was clear because the corpus was self-created. As illustrated in the method testing in Study 1, setting the correct number of clusters to identify topics is a challenge when information about the expected number of clusters is unavailable. Therefore, the subjectivity of determining the number of clusters for text mining was countered in Study 2, in which the participating students were instructed to describe one topic per paragraph of their essays. Based on this strategy, the average number of paragraphs was assumed as a suitable cluster number for topic identification. However, some students have forgotten to divide their text into paragraphs; hence, this rule for determining the number of clusters did not turn out to be optimal. Furthermore, varying this number to make a final decision on the meaningful representation of the topics from the learners' essays (as undertaken with LDA in Study 1) also lacks objectivity. Similarly in the procedure of specifying content, the instructors' determination of frequency thresholds and interpretation of concept clusters to identify topics limit the objectivity. This limitation can impair the accuracy of the provided information about the content, the importance of which has been explained in detail (see section 1.1.). It might also be decisive for how effectively the specification of information about learning content activates prior knowledge. Missing accuracy should therefore be prevented to avoid the impairment of knowledge exchange.

In addition to these aspects of text mining-based grouping and representing, guidance could support better focusing topics to be learned when representing content. The exploratory observations in Study 2 lead to the conclusion that precisely such topics that are known to only one or no learning partner in a learning group seem to be addressed in the knowledge exchange. However, when selecting from several listed topics with such knowledge distribution, learners seem to pick topics that are more associated with everyday knowledge than with facts. This observation underlines that in addition to their accurate provision, these expert topics should not be deprived of consideration in a bottom-up approach (if learner-generated texts, for example, primarily report on everyday knowledge). Furthermore, taking into account the results from Study 3 indicating on the one hand that prior knowledge is not better activated by the text mining-based bottom-up approach, and on the other hand that provided content can direct questions, a top-down approach might be the better choice. The listing of previously defined topics, which are considered important by the instructor and to be the focus in knowledge exchange, could be supported by a concept classification automated by text-mining methods and be based on a pre-determinable number of classes to overcome the aforementioned problem.

Study 3 also has some limitations, particularly in the operationalization of cognitive processing and its distinction from communication. Regarding cognitive processing, it may be error-prone to infer cognitive processes from verbal communication, such as the activation of prior knowledge from the mentioned number of task-relevant topics or the level of cognitive elaboration from the number of integrated topic explanations; both may also have occurred independently from verbalizing it, and a clearer distinction should be made between cognitive processing and communication to gain an increased understanding of how both are related in the context of implicit guidance. Matching the previous limitation is another one that is associated with communication that, in favor of a systematic examination, was based only on a simulated collaboration. Learning can only transpire from communicating with others when learners receive explanations in response to their questions or offer explanations in response to questions. Hence, the influence of arising communication strategies on learning outcomes could therefore not be considered in this work because the communication was only simulated and might be different in a real collaborative scenario. For this reason, the effect of both types of information on verbal exchange in a real collaboration scenario and its influence on learning outcomes should still be investigated systematically. A description of further limitations is available in the respective articles. The limitations presented in this section have been identified according to their relevance for the recommendations for further research.

1.5.3. Recommendations for further research

Instructors need to guide cognitive processing and communication in knowledge exchange so that people can learn. The overall objective of this work is to explore the potential of text-mining methods to facilitate the instructors' work and improve knowledge exchange in collaborative learning, with the results of this work also informing tool designs. The aforementioned limitations hindering the achievement of these goals and the unexpected results from the three studies indicate some aspects that would be interesting for further investigating the functions of text-mining methods (see Figure 4, p. 19). In addition, the paths of the model of guided knowledge exchange (see Figure 3, p. 13) that have not been examined in this work might suggest interesting directions for future research to complement the understanding of the mechanisms triggered by instructional guidance.

Regarding text mining-based grouping and representing, the main tasks of method-based future studies should be the optimization of the grouping algorithm, further exploring the question of the specificity of visualized information about the content and varying the processing of information about learning content. As described in the limitations (see section 1.5.2.), the grouping requires optimization. Thus, a new grouping algorithm should be tested which, for example, sets a distance value as a threshold value above which composed learning partners must lie in order to be grouped. In the case of clear differences between random and algorithm-based formations of learning partners, the use of text mining-based groupings would then be optimized for usage in smaller overall groups (e.g., in classroom settings). Another research direction would be the examination of whether the determination of heterogeneity can be extended or adapted to better capture complementarity (e.g., by prescribing the text structure, determining Euclidean distance paragraph by paragraph and including the claim of an even distribution of cases of expertise) or to even identify diverse opinions (e.g., based on words or phrases signaling that a person adopts a different view). Some conceptual work and comparison of approaches would be interesting.

Aside from optimizing the grouping, the question of the specificity of the provided information about the learning content should be investigated further. As described in the discussion (see section 1.5.1.), the number of listed topics and concepts might be too high and cause a cognitive overload, which renders the necessity of reducing the number of topics, the number of concepts, or both. From a scientific viewpoint, a noteworthy research direction would be ascertaining how far this reduction would have to go to provide learners with optimal support. However, from a methodological perspective, this suggested research direction raises the question on the extent to which a reduction in the number of clusters influences the accuracy of text-mining methods (topics may be lost). Moreover, the question of the number of topics and concepts that must be displayed to accurately specify the learning content and trigger the corresponding communication emerges (see limitations in section 1.5.2.). To answer these questions, future research should systematically decrease the number of visualized concepts and related topics (compared to nine topics in Study 2 and eight topics in Study 3). On the one hand, exploring how effectively text mining could still accurately identify relevant topics would be of interest. On the other hand, determining the appropriate number of topics and concepts to allow the learners a comprehensible overview of the important topics for knowledge exchange would be relevant.

However, the text mining-based processing of information about the content should be varied. As described in the limitations section (see section 1.5.2.), text mining-based methods, which do not adopt a bottom-up approach (corresponding to concept clustering) but a top-down approach (corresponding to concept classification), might improve the focus on relevant topics in knowledge exchange by providing the information resulting from this alternative analysis. To examine this assumption, text mining-based classifications would merit further exploration. For example, this exploration could be based on training classification algorithms with written learning material and applying them to learner-generated texts, thereby determining the degree to which or how effectively each learner-generated text reflects the topics of the learning material. On the one hand, verifying the most suitable classification method for evaluating the quality of a learner-generated text would be of interest. On the other hand, a relevant focus of future research would be determining whether the provision of classification-based information about content would achieve other guidance effects on communication and learning compared to the provision of information collected by concept clustering.

The further investigation of mechanisms can help to improve instructional guidance. Figure 10 (p. 51) illustrates the paths in the model that would be particularly interesting. (1) To enhance the understanding of the added value of complementarity-based grouping in implicit guidance, its influence on learning, alone and in combination with group awareness support, should be further analyzed as both were confounded in Study 2 (see the discussion in section 1.5.1.). (2) Moreover, communication in the present work was merely simulated, and it could vary in real collaborations (see limitations in section 1.5.2.). For this reason, the effect of both types of information on verbal exchange in a real collaboration scenario and the influence of this exchange on learning still require a systematic investigation. (3) Along with new insights into this aspect, a noteworthy direction would be understanding how communication influences cognitive processing—or the other way around—and how they interact with each other. For instance, future studies could examine how groups differ when one group only cognitively elaborates on what would be explained to the partner on certain

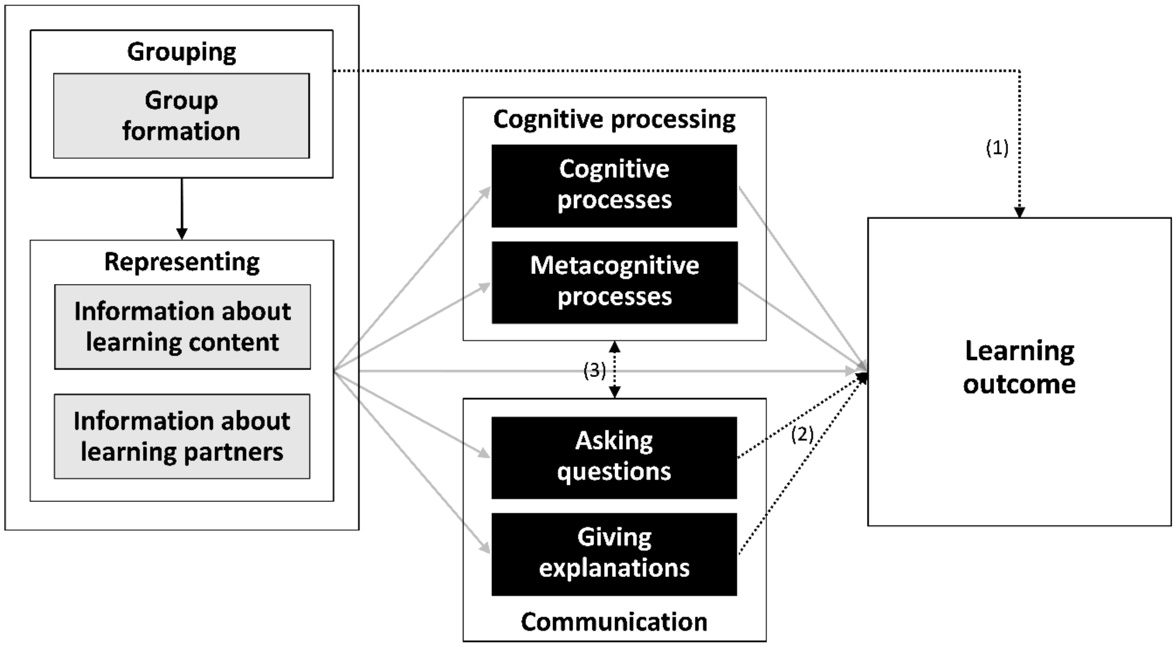

Figure 10. Possible foci of analyses in future studies. The dotted lines denote the paths that remain uninvestigated in the model or would be of interest after an adaptation of methods.

topics, while the other group additionally exchanges information. If the provision of various types of information is thereby considered, information about the contradictory results related to the content specificity of information could be provided as well.

1.5.4. Conclusion

The increasing digitalization of education, workplaces, and private life requires many skills to design social interactions in digital learning environments; at the same time, it creates innovative means of automating instructions for interaction support. Against this background, the results of the current work constitute a foundation for increasing the adoption of text-mining methods for further exploiting their potential to boost the efficiency of guidance in these environments and promote learning. This thesis sets a general precedent for the use of digital text artifacts for automatically collecting cognitive information about people and for transforming and representing it to guide their knowledge exchange. Specifically, this work proposes an innovative model of how group formation and group awareness support can be combined and offers answers on how this model can be applied through the support of text-mining methods. Moreover, new insights into mechanisms triggered by the provided cognitive information, whether they are collected by text-mining methods or not, can enrich other areas of (technology-enhanced) learning, thereby facilitating the instructors' work and improving cognitive processing and communication.

2. Paper 1: Improving collaborative learning in the classroom: Text mining-based grouping and representing (Studies 1 and 2)

Erkens, M., Bodemer, D., & Hoppe, H. U. (2016). Improving collaborative learning in the classroom: Text mining based grouping and representing. *International Journal of Computer-Supported Collaborative Learning, 11*(4), 387–415. doi:10.1007/s11412-016-9243-5

Retrievable from https://link.springer.com/article/10.1007/s11412-016-9243-5

3. Paper 2: Which visualization guides learners best? Impact of available partner- and content-related information on collaborative learning (Study 3)

Erkens, M., & Bodemer, D. (2017). Which visualization guides learners best? Impact of available partner- and content-related information on collaborative learning. In B. K. Smith, M. Borge, E. Mercier, & K. Y. Lim (Eds.), *Making a Difference: Prioritizing Equity and Access in CSCL, 12th International Conference on Computer Supported Collaborative Learning (CSCL) 2017* (Vol. 1, pp. 127–134). Philadelphia, PA: International Society of the Learning Sciences.

Retrievable from https://repository.isls.org/bitstream/1/223/1/20.pdf

4. Paper 3: Improving collaborative learning: Guiding knowledge exchange through the provision of information about learning partners and learning contents (Study 3)

Erkens, M., & Bodemer, D. (2019). Improving collaborative learning: Guiding knowledge exchange through the provision of information about learning partners and learning contents. *Computers & Education, 128*(1), 452–472. doi:10.1016/j.compedu.2018.10.009

Retrievable from
https://www.sciencedirect.com/science/article/pii/S036013151830280X?via%3Dihub

5. Bibliography

Aggarwal, C. C., & Zhai, C. (2012). A survey of text clustering algorithms. In C. C. Aggarwal & C. Zhai (Eds.), *Mining text data* (pp. 77–128). New York, NY: Springer.
Ames, G. J., & Murray, F. B. (1982). When two wrongs make a right: Promoting cognitive change by social conflict. *Developmental Psychology*, *18*(6), 894–897. doi:10.1037/0012-1649.18.6.894
Aronson, E., Blaney, N., Stephan, C., Sikes, J., & Snapp, M. (1978). *The jigsaw classroom*. Beverly Hills, CA: Sage.
Ausubel, D. P., Novak, J. D., & Hanesian, H. (1978). *Educational psychology: A cognitive view* (2nd ed.). New York, NY: Holt, Rinehart and Winston.
Ausubel, D. P. (2000). Assimilation theory in meaningful learning and retention processes. In D. P. Ausubel (Eds.), *The acquisition and retention of knowledge: A cognitive view* (pp. 101–145). Dordrecht, Netherlands: Springer.
Bandura, A., & Jourden, F. J. (1991). Self-regulatory mechanisms governing the impact of social comparison on complex decision making. *Journal of Personality and Social Psychology*, *60*(6), 941–951. doi:10.1037/0022-3514.60.6.941
Baron, J. (1991). Beliefs about thinking. In J. F. Voss, D. N. Perkins, & J. W. Segal (Eds.), *Informal reasoning and education* (pp. 169–186). Hillsdale, NJ: Lawrence Erlbaum.
Berger, A., Moretti, R., Chastonay, P., Dillenbourg, P., Bchir, A., Baddoura, R. M., … Kayser, B. (2001). Teaching community health by exploiting international socio-cultural and economical differences. In P. Dillenbourg, A. Eurelings & K. Hakkarainen (Eds.), *European Perspectives on Computer-supported Collaborative Learning: Proceedings of the First European Conference on Computer-supported Collaborative Learning* (pp. 97–105). Maastricht, Netherlands: Euro-CSCL. Retrieved from http://www.eculturenet.org/mmi/euro-cscl/Papers/14.pdf
Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM, 55*(4), 77–84. doi:10.1145/2133806.2133826
Bodemer, D. (2011). Tacit guidance for collaborative multimedia learning. *Computers in Human Behavior*, *27*(3), 1079–1086. doi:10.1016/j.chb.2010.05.016
Bodemer, D., & Buder, J. (2006). Supporting collaborative learning with augmented group awareness tools. In R. Sun & N. Miyake (Eds.), *Proceedings of the Twenty-Eighth Annual Conference of the Cognitive Science Society* (pp. 77–82). Mahwah, NJ: Lawrence Erlbaum. Retrieved from https://escholarship.org/uc/item/0rn2v7cj
Bodemer, D., & Dehler, J. (2011). Group awareness in CSCL environments. *Computers in Human Behavior*, *27*(3), 1043–1045. doi:10.1016/j.chb.2010.07.014
Bodemer, D., Janssen, J., & Schnaubert, L. (2018). Group awareness tools for computer-supported collaborative learning. In F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, & P. Reimann (Eds.), *International Handbook of the Learning Sciences* (pp. 351–358). New York, NY: Routledge / Taylor & Francis.
Bodemer, D., & Scholvien, A. (2014). Providing knowledge-related partner information in collaborative multimedia learning: Isolating the core of cognitive group awareness tools. In C.-C. Liu, H. Ogata, S. C. Kong, & A. Kashihara (Eds.), *Proceedings of the 22nd International Conference on Computers in Education ICCE 2014* (pp. 171–179). Nomi, Japan: ICCE 2014 Organizing Committee / ICT Unit, Center for Graduate Education Initiative, Japan Advanced Institute of Science and Technology. Retrieved from http://www.apsce.net/icce/icce2014/140.115.135.84/icce/icce2014/wp-content/ uploads/2014/12/ICCE2014-Main-Proceedings-lite-1.pdf
Buder, J., & Bodemer, D. (2008). Supporting controversial CSCL discussions with augmented group awareness tools. *International Journal of Computer-Supported Collaborative Learning*, *3*(2), 123–139. doi:10.1007/s11412-008-9037-5

Buder, J. (2017). A conceptual framework of knowledge exchange. In S. Schwan & U. Cress (Eds.), *The Psychology of Digital Learning* (pp. 105–122). Cham, Switzerland: Springer.

Callender, A. A. (2012). Learning from questions. In N. M. Seel (Ed.), *Encyclopedia of the Sciences of Learning* (1st ed., pp. 1867–1869). Boston, MA: Springer.

Clark, H. H., & Murphy, G. L. (1982). Audience design in meaning and reference. *Advances in Psychology*, *9*(C), 287–299. doi:10.1016/S0166-4115(09)60059-5

Crommelinck, M., & Anseel, F. (2013). Understanding and encouraging feedback-seeking behaviour: A literature review. *Medical Education, 47*(3), 232–241. doi:10.1111/medu.12075

Crossley, S. A., Allen, L. K., Snow, E. L., & McNamara, D. S. (2015). Pssst... textual features... There is more to automatic essay scoring than just you! In J. Baron, G. Lynch, N. Maziarz, P. Blikstein, A. Merceron, & G. Siemens (Eds.), *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge* (pp. 203–207). Poughkeepsie, NY: ACM. doi:10.1145/2723576.2723595

Daems, O., Erkens, M., Malzahn, N., & Hoppe, H. U. (2014). Using content analysis and domain ontologies to check learners' understanding of science concepts. *Journal of Computers in Education*, *1*(2–3), 113–131. doi:10.1007/s40692-014-0013-y

Dehler, J., Bodemer, D., Buder, J., & Hesse, F. W. (2009). Providing group knowledge awareness in computer-supported collaborative learning: Insights into learning mechanisms. *Research and Practice in Technology Enhanced Learning*, *4*(2), 111–132. doi:10.1142/S1793206809000660

Dehler, J., Bodemer, D., Buder, J., & Hesse, F. W. (2011). Guiding knowledge communication in CSCL via group knowledge awareness. *Computers in Human Behavior*, *27*(3), 1068–1078. doi:10.1016/j.chb.2010.05.018

Dehler Zufferey, J., Bodemer, D., Buder, J., & Hesse, F. W. (2011). Partner knowledge awareness in knowledge communication: Learning by adapting to the partner. *The Journal of Experimental Education*, *79*(1), 102–125. doi:10.1080/00220970903292991

Deiglmayr, A., & Schalk, L. (2015). Weak versus strong knowledge interdependence: A comparison of two rationales for distributing information among learners in collaborative learning settings. *Learning and Instruction, 40*, 69–78. doi:10.1016/j.learninstruc.2015.08.003

Dillenbourg, P. (1999). What do you mean by "collaborative learning"? In P. Dillenbourg (Ed.), *Collaborative Learning: Cognitive and Computational Approaches* (pp. 1–19). Oxford, England: Elsevier.

Dillenbourg, P., & Jermann, P. (2007). Designing integrative scripts. In F. Fischer, I. Kollar, H. Mandl, & J. M. Haake (Eds.), *Scripting Computer-Supported Collaborative Learning* (pp. 275–301). New York, NY: Springer.

Dillenbourg, P. (2002). Over-scripting CSCL: The risks of blending collaborative learning with instructional design. In P. A. Kirschner (Ed.), *Three worlds of CSCL. Can we support CSCL?* (pp. 61–91). Heerlen, Netherlands: Open Universiteit Nederland.

Dillenbourg, P., & Bétrancourt, M. (2006). Collaboration load. In J. Elen & R. E. Clark (Eds.), *Handling complexity in learning environments: Research and theory* (pp. 142–163). Amsterdam, Netherlands: Elsevier.

Dochy, F., de Rijdt, C., & Dyck, W. (2002). Cognitive prerequisites and learning. *Active Learning in Higher Education*, *3*(3), 265–284. doi:10.1177/1469787402003003006

Dochy, F., Segers, M., & Buehl, M. M. (1999). The relation between assessment practices and outcomes of studies: The case of research on prior knowledge. *Review of Educational Research*, *69*(2), 145–186. doi:10.3102/00346543069002145

Doise, W., & Hanselmann, C. (1991). Conflict and social marking in the acquisition of operational thinking. *Learning and Instruction*, *1*(2), 119–127. doi:10.1016/0959-4752(91)90022-Z

Doise, W., & Mugny, G. (1984). *The social development of the intellect*. Oxford, England: Pergamon Press.

Dufty, D. F., Graesser, A. C., Louwerse, M. M., & McNamara, D. S. (2006). Assigning grade levels to textbooks: Is it just readability? In R. Sun & N. Miyake (Eds.), *Proceedings of the Annual Meeting of the Cognitive Science Society 28* (pp. 1251–1256). Austin, TX: Cognitive Science Society. Retrieved from https://escholarship.org/uc/item/44f184z

Dunlosky, J., & Metcalfe, J. (2009). *Metacognition*. Los Angeles, CA: Sage.

Efklides, A. (2008). Metacognition: Defining its facets and levels of functioning in relation to self-regulation and co-regulation. *European Psychologist*, *13*(4), 277–287. doi:10.1027/1016-9040.13.4.277

Erkens, M., & Bodemer, D. (2017). Which visualization guides learners best? Impact of available partner- and content-related information on collaborative learning. In B. K. Smith, M. Borge, E. Mercier, & K. Y. Lim (Eds.), *Making a Difference: Prioritizing Equity and Access in CSCL, 12th International Conference on Computer Supported Collaborative Learning (CSCL) 2017* (Vol. 1, pp. 127–134). Philadelphia, PA: International Society of the Learning Sciences. Retrieved from https://repository.isls.org/bitstream/1/223/1/20.pdf

Erkens, M., Daems, O., & Hoppe, H. U. (2014). Artifact analysis around video creation in collaborative STEM learning scenarios. In L. O'Conner (Ed.), *2014 IEEE 14th International Conference on Advanced Learning Technologies (ICALT)* (pp. 388–392). Athens, Greece. doi:10.1109/ICALT.2014.116

Erkens, M., & Bodemer, D. (2019). Improving collaborative learning: Guiding knowledge exchange through the provision of information about learning partners and learning contents. *Computers & Education*, *128*, 452–472. doi:10.1016/j.compedu.2018.10.009

Erkens, M., Bodemer, D., & Hoppe, H. U. (2016). Improving collaborative learning in the classroom: Text mining based grouping and representing. *International Journal of Computer-Supported Collaborative Learning*, *11*(4), 387–415. doi:10.1007/s11412-016-9243-5

Fawcett, L. M., & Garton, A. F. (2005). The effect of peer collaboration on children's problem-solving ability. *The British Journal of Educational Psychology*, *75*(2), 157–169. doi:10.1348/000709904X23411

Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, *4*(5/6), 304–317. doi:10.1504/IJTEL.2012.051816

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, *7*(2), 117–140. doi:10.1177/001872675400700202

Fischer, F., Kollar, I., Stegmann, K., & Wecker, C. (2013). Toward a script theory of guidance in computer-supported collaborative learning. *Educational Psychologist*, *48*(1), 56–66. doi:10.1080/00461520.2012.748005

Foltz, P. W., & Rosenstein, M. (2015). Analysis of a large-scale formative writing assessment system with automated feedback. *Proceedings of the Second ACM Conference on Learning @ Scale*, 339–342. doi:10.1145/2724660.2728688

Gokhale, A. A. (1995). Collaborative learning enhances critical thinking. *Journal of Technology Education*, *7*(1), 22–30. doi:10.1007/978-1-4419-1428-6_910

Gurlitt, J., Dummel, S., Schuster, S., & Nückles, M. (2011). Differently structured advance organizers lead to different initial schemata and learning outcomes. *Instructional Science, 40*(2), 351–369. doi:10.1007/s11251-011-9180-7.

Hare, A. P. (1994). Types of roles in small groups: A bit of history and a current perspective. *Small Group Research, 25*(3), 433–448. doi:1046496494253005

Hesse, F. (2007). Being told to do something or just being aware of something? An alternative approach to scripting in CSCL. In F. Fischer, I. Kollar, H. Mandl, & J. M. Haake (Eds.), *Scripting Computer-Supported Communication: Cognitive, Computational and Educational Perspectives* (pp. 91–98). New York, NY: Springer.

Hoppe, H. U. (2017). Computational methods for the analysis of learning and knowledge building communities. In C. Lang, G. Siemens, A. Wise, & D. Gašević (Eds.), *Handbook of Learning Analytics* (pp. 23–33). Beaumont, Canada: Society for Learning Analytics Research. doi:10.18608/hla17

Hotho, A., Nürnberger, A., & Paaß, G. (2005). A brief survey of text mining. *GLDV Journal for Computational Linguistics and Language Technology, 20*(1), 19–62.

Jacobs, G. M., Power, M. A., & Loh, W. I. (2016). *The Teacher's Sourcebook for Cooperative Learning: Practical Techniques, Basic Principles, and Frequently Asked Questions*. New York, NY: Skyhorse.

Janssen, J., & Bodemer, D. (2013). Coordinated computer-supported collaborative learning: Awareness and awareness tools. *Educational Psychologist*, *48*(1), 40–55. doi:10.1080/00461520.2012.749153

Johnson, D. W., Johnson, R. T., & Stanne, M. B. (2000). *Cooperative learning methods: A meta-analysis* (pp. 1–33). Minneapolis, MN: University of Minnesota. Retrieved from https://www.researchgate.net/publication/220040324_Cooperative_Learning_Methods_A_Meta-Analysis

Johnson, D. W., & Johnson, R. T. (1979). Conflict in the classroom: Controversy and learning. *Review of Educational Research, 49*(1), 51–69. doi:10.3102/00346543049001051

Johnson, D. W., & Johnson, R. T. (2009). Energizing learning: The instructional power of conflict. *Educational Researcher*, *38*(1), 37–51. doi:10.3102/0013189X08330540

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist, 58*(9), 697–720. doi:10.1037/0003-066X.58.9.697

Kalyuga, S. (2009). Knowledge elaboration: A cognitive load perspective. *Learning and Instruction*, *19*(5), 402–410. doi:10.1016/j.learninstruc.2009.02.003

King, A. (1994). Guiding knowledge construction in the classroom: Effects of teaching children how to question and how to explain. *American Educational Research Journal*, *31*(2), 338–368. doi:10.3102/00028312031002338

Knight, S., & Buckingham Shum, S. (2017). Theory and learning analytics. In C. Lang, G. Siemens, A. Wise, & D. Gašević (Eds.), *Handbook of Learning Analytics* (pp. 17–22). Beaumont, Canada: Society for Learning Analytics Research. doi:10.18608/hla17

Kobbe, L., Weinberger, A., Dillenbourg, P., Harrer, A., Hämäläinen, R., Häkkinen, P., & Fischer, F. (2007). Specifying computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning*, *2*(2), 211–224. doi:10.1007/s11412-007-9014-4

Kollar, I., Fischer, F., & Hesse, F. W. (2006). Collaboration scripts – a conceptual analysis. *Educational Psychology Review*, *18*(2), 159–185. doi:10.1007/s10648-006-9007-2

Koriat, A. (2007). Metacognition and consciousness. In P.D. Ze1azo, M. Moscovitch, & E. Thompson (Eds.), *The Cambridge handbook of consciousness* (pp. 289–325). Cambridge, England: Cambridge University Press.

Kovanović, V., Joksimović, S., Gašević, D., Hatala, M., & Siemens, G. (2017). Content analytics: The definition, scope, and an overview of published research. In C. Lang, G. Siemens, A. Wise, & D. Gašević (Eds.), *Handbook of Learning Analytics* (pp. 77–92). Beaumont, Canada: Society for Learning Analytics Research. doi:10.18608/hla17

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, *77*(6), 1121–1134. doi:10.1037/0022-3514.77.6.1121

Lin, L. (2015). Exploring collaborative learning: Theoretical and conceptual perspectives. In L. Lin (Ed.), *Investigating Chinese HE EFL Classrooms: Using Collaborative Learning to Enhance Learning* (pp. 11–28). Berlin / Heidelberg, Germany: Springer. doi:10.1007/978-3-662-44503-7_2

Linn, M. C. (2005). The knowledge integration perspective on learning and instruction. In R. K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (pp. 243–264). Cambridge, England: Cambridge University Press. doi:10.1017/CBO9780511816833. 016

Lipponen, L. (2002). Exploring Foundations for Computer-supported Collaborative Learning. In G. Stahl (Ed.), *Proceedings of the Conference on Computer Support for Collaborative Learning: Foundations for a CSCL Community* (pp. 72–81). Boulder, CO: International Society of the Learning Sciences. Retrieved from http://dl.acm.org/citation.cfm?id=1658616.1658627

Liu, R., & Koedinger, K. R. (2015). Variations in learning rate: Student classification based on systematic residual error patterns across practice opportunities. In O. C. Santos et al. (Eds.), *Proceedings of the 8th International Conference on Education Data Mining (EDM2015)*. Worcester, MA: Educational Data Mining Society. Retrieved from https://files.eric.ed.gov/fulltext/ED560874.pdf

Lockridge, C. B., & Brennan, S. E. (2002). Addressees' needs influence speakers' early syntactic choices. *Psychonomic Bulletin & Review*, *9*(3), 550–557. doi:10.3758/BF03196312

Mäkitalo-Siegl, K., & Kollar, I. (2012). Collaboration scripts. In N. M. Seel (Ed.), *Encyclopedia of the sciences of learning* (pp. 628–631). New York, NY: Springer.

Mayer, R. E. (1979). *Twenty years of research on advance organizers.* (Technical Report No. 79-1) Retrieved from https://files.eric.ed.gov/fulltext/ED206691.pdf

Mayer, R. E. (2001). A cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *Multimedia learning* (pp. 41–62). Cambridge, England: Cambridge University Press.

Mayer, R. E. (2010). Merlin C. Wittrock's enduring contributions to the science of learning. *Educational Psychologist*, *45*(1), 46–50. doi:10.1080/00461520903433547

Merriënboer, J. J. G. van, & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review, 17*(2), 147–177. doi:10.1007/s10648-005-3951-0

Miner, G. D., Elder, J., Fast, A., Hill, T., Nisbet, B., & Delen, D. (2012). *Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications*. Boston, MA: Academic Press.

Moreno, R., & Mayer, R. (2007). Interactive multimodal learning environments: Special issue on interactive learning environments: Contemporary issues and trends. *Educational Psychology Review*, *19*(3), 309–326. doi:10.1007/s10648-007-9047-2

Nelson, T. O. (1996). Consciousness and metacognition. American Psychologist, 51(2), 102-116. doi:10.1037/0003-066X.51.2.102

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *Psychology of learning & motivation* (Vol. 26, pp. 125–173). New York, NY: Academic Press.

Nelson, T. O., & Narens, L. (1994). Why investigate metacognition? In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 1–25). Cambridge, MA: Bradford Books.

Neugebauer, J., Ray, D. G., & Sassenberg, K. (2016). When being worse helps: The influence of upward social comparisons and knowledge awareness on learner engagement and learning in peer-to-peer knowledge exchange. *Learning and Instruction*, *44*, 41–52. doi:10.1016/j.learninstruc.2016.02.007

Nickerson, R. S. (1999). How we know—and sometimes misjudge—what others know: Imputing one's own knowledge to others. *Psychological Bulletin*, *125*(6), 737–759. doi:10.1037/0033-2909.125.6.737

Nugent, R., Ayers, E., & Dean, N. (2009). Conditional Subspace Clustering of Skill Mastery: Identifying Skills that Separate Students. In T. Barnes, M. Desmarais, C. Romero, & S. Ventura (Eds.), *Educational Data Mining 2009* (pp. 101-110), Cordoba, Spain: International Educational Data Mining Society. Retrieved from https://files.eric.ed.gov/fulltext/ED539073.pdf

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

Perera, D., Kay, J., Koprinska, I., Yacef, K., & Zaïane, O. R. (2009). Clustering and sequential pattern mining of online collaborative learning data. *IEEE Transactions on Knowledge and Data Engineering*, *21*(6), 759–772. doi:10.1109/TKDE.2008.138

Piaget, J. (1959). *The language and thought of the child* (3rd ed.). New York, NY: Routledge / Taylor & Francis.

Piaget, J. (1977). *The development of thought: equilibration of cognitive structures*. New York, NY: Viking Press.

Ploetzner, R., Dillenbourg, P., Preier, M., & Traum, D. (1999). Learning by explaining to oneself and to others. In P. Dillenbourg (Ed.), *Collaborative Learning: Cognitive and Computational Approaches* (pp. 103–121). Oxford, England: Elsevier.

Ray, D. G., Neugebauer, J., Sassenberg, K., Buder, J., & Hesse, F. W. (2013). Motivated shortcomings in explanation: the role of comparative self-evaluation and awareness of explanation recipient's knowledge. *Journal of Experimental Psychology: General*, *142*(2), 445–457. doi:10.1037/a0029339

Reimann, P. P. (2018). Making online group-work work: Scripts, group awareness and facilitation. In G. Masters, J. Ainley, D. Buehring, H. Hollingsworth, M. Hughes, L. Ingvarson, …. S. Thomson, *Research Conference 2008. Teaching practices that make a difference: Insights from Research* (pp. 49–52). Melbourne, Australia: Australian Council for Educational Research. Retrieved from https://research.acer.edu.au/cgi/viewcontent.cgi?article=1338&context=research_conference

Romero, M., & Lambropoulos, N. (2017). Internal and external regulation to support knowledge construction and convergence in computer supported collaborative learning (CSCL). *Electronic Journal of Research in Education Psychology*, *9*(1), 309–330. doi:10.25115/ejrep.v9i23.1439

Roschelle, J., & Teasley, S. D. (1995). The construction of shared knowledge in collaborative problem solving. In C. O'Malley (Ed.), *Computer Supported Collaborative Learning* (pp. 69–97). Berlin, Germany: Springer.

Salonen, P., Vauras, M., & Efklides, A. (2005). Social interaction - what can it tell us about metacognition and coregulation in learning? *European Psychologist, 10*(3), 199–208. doi:10.1027/1016-9040.10.3.199

Sangin, M., Molinari, G., Nüssli, M.-A., & Dillenbourg, P. (2008). Knowing what the peer knows: The differential effect of knowledge awareness on collaborative learning performance of asymmetric pairs. In P. Dillenbourg & M. Specht (Eds.), *Times of Convergence. Technologies Across Learning Contexts* (pp. 384–394). Berlin, Germany: Springer. doi:10.1007/978-3-540-87605-2_43

Sangin, M., Molinari, G., Nüssli, M.-A., & Dillenbourg, P. (2011). Facilitating peer knowledge modeling: Effects of a knowledge awareness tool on collaborative learning outcomes and processes. *Computers in Human Behavior*, *27*(3), 1059–1067. doi:10.1016/j.chb.2010.05.032

Scardamalia, M., & Bereiter, C. (1994). Computer support for knowledge-building communities. *Journal of the Learning Sciences*, *3*(3), 265–283. doi:10.1207/s15327809jls0303_3

Scardamalia, M., & Bereiter, C. (2014). Smart technology for self-organizing processes. *Smart Learning Environments*, *1*(1), 1-13. doi:10.1186/s40561-014-0001-8

Schittekatte, M., & Hiel, A. V. (1996). Effects of partially shared information and awareness of unshared information on information sampling. *Small Group Research, 27*(3), 431–449. doi:10.1177/1046496496273006

Schneider, W., Körkel, J., & Weinert, F. E. (1989). Domain-specific knowledge and memory performance: A comparison of high- and low-aptitude children. *Journal of Educational Psychology*, *81*(3), 306–312. doi:10.1037/0022-0663.81.3.306

Schnotz, W., & Preuß, A. (1999). Task-dependent construction of mental models as a basis for conceptual change. In W Schnotz, S. Vosniadou, & M. Carretero (Eds.), *New perspectives on conceptual change* (pp. 193–222). Oxford, England: Pergamon Press.

Schober, M. F., & Brennan, S. E. (2003). Processes of interactive spoken discourse: The role of the partner. In A. C. Graesser, M. A. Gernsbacher, & S. R. Goldman (Eds.), *Handbook of Discourse Processes* (pp. 123–164). Mahwah, NJ: Lawrence Erlbaum.

Sherin, B. (2012). Using computational methods to discover student science conceptions in interview data. In S. Buckingham Shum, D. Gašević, & R. Ferguson (Eds.), *Proceedings of the 2Nd International Conference on Learning Analytics and Knowledge* (pp. 188–197). New York, NY: ACM. doi:10.1145/2330601.2330649

Southavilay, V., Yacef, K., Reimann, P., & Calvo, R. A. (2013). Analysis of collaborative writing processes using revision maps and probabilistic topic models. In D. Suthers, K. Verbert, E. Duval, & X. Ochoa (Eds.), *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp. 38–47). Leuven, Belgium: ACM. doi:10.1145/2460296.2460307

Stark, R., Mandl, H., Gruber, H., & Renkl, A. (2002). Conditions and effects of example elaboration. *Learning and Instruction*, *12*(1), 39–60. doi:10.1016/S0959-4752(01)00015-9

Stegmann, K., Wecker, C., Weinberger, A., & Fischer, F. (2012). Collaborative argumentation and cognitive elaboration in a computer-supported collaborative learning environment. *Instructional Science*, *40*(2), 297–323. doi:10.1007/s11251-011-9174-5

Tansomboon, C., Gerard, L. F., Vitale, J. M., & Linn, M. C. (2017). Designing automated guidance to promote productive revision of science explanations. *International Journal of Artificial Intelligence in Education*, *27*(4), 729–757. doi:10.1007/s40593-017-0145-0

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. In Cole, M., John-Steiner, V., Scribner, S., & Souberman, E. (Eds.). *Mind in society: The development of higher psychological processes* (Revised ed. edition). Cambridge, MA: Harvard University Press.

Webb, N. M. (1989). Peer interaction and learning in small groups. *International Journal of Educational Research*, *13*(1), 21–39. doi:10.1016/0883-0355(89)90014-1

Wecker, C., & Fischer, F. (2014). Lernen in Gruppen [Group learning]. In T. Seidel & A. Krapp (Eds.), *Pädagogische Psychologie* [*Educational Psychology*] (pp. 277–296). Weinheim, Germany: Beltz.

Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education*, *46*(1), 71–95. doi:10.1016/j.compedu.2005.04.003

Weinberger, A., Stegmann, K., & Fischer, F. (2010). Learning to argue online: Scripted groups surpass individuals (unscripted groups do not). *Computers in Human Behavior*, *26*(4), 506–515. doi:10.1016/j.chb.2009.08.007

Weinberger, A., Stegmann, K., Fischer, F., & Mandl, H. (2007). Scripting argumentative knowledge construction in computer-supported learning environments. In F. Fischer, I. Kollar, H. Mandl, & J. M. Haake (Eds.), *Scripting computer-supported collaborative learning* (pp. 191–211). New York, NY: Springer.

Weinert, F. E., & Helmke, A. (1998). The neglected role of individual differences in theoretical models of cognitive development. *Learning and Instruction*, *8*(4), 309–323. doi:10.1016/S0959-4752(97)00024-8

DOI: 10.17185/duepublico/70177
URN: urn:nbn:de:hbz:464-20190625-091437-6