CrossMark

Improving collaborative learning in the classroom: Text mining based grouping and representing

Melanie Erkens[1] (iD) · Daniel Bodemer[1] ·
H. Ulrich Hoppe[2]

Abstract Orchestrating collaborative learning in the classroom involves tasks such as forming learning groups with heterogeneous knowledge and making learners aware of the knowledge differences. However, gathering information on which the formation of appropriate groups and the creation of graphical knowledge representations can be based is very effortful for teachers. Tools supporting cognitive group awareness provide such representations to guide students during their collaboration, but mainly rely on specifically created input. Our work is guided by the questions of how the analysis and visualization of cognitive information can be supported by automatic mechanisms (especially using text mining), and what effects a corresponding tool can achieve in the classroom. We systematically compared different methods to be used in a *Grouping and Representing Tool* (GRT), and evaluated the tool in an experimental field study. Latent Dirichlet Allocation proved successful in transforming the topics of texts into values as a basis for representing cognitive information graphically. The Vector Space Model with Euclidian distance based clustering proved to be particularly well suited for detecting text differences as a basis for group formation. The subsequent evaluation of the GRT with 54 high school students further confirmed the GRT's impact on learning support: students who used the tool added twice as many concepts in an essay after discussing as those in the unsupported group. These results show the potential of the GRT to support both teachers and students.

✉ Melanie Erkens
   melanie.erkens@uni-due.de

✉ Daniel Bodemer
   bodemer@uni-due.de

   H. Ulrich Hoppe
   hoppe@collide.info

[1] University of Duisburg-Essen (Media-Based Knowledge Construction), Lotharstr. 65,
     47057 Duisburg, Germany

[2] University of Duisburg-Essen (Collaborative Learning in Intelligent Distributed Environments),
     Lotharstr. 63, 47057 Duisburg, Germany

⚫ Springer

Introduction

With the growing popularity of educational technologies, teachers are increasingly required to orchestrate their classrooms across different contexts, media, and multiple learning activities (Dillenbourg et al. 2009). This is particularly true if they apply pedagogical techniques such as collaborative learning (e.g. Dimitriadis 2012). Since learners frequently do not engage in collaborative learning activities on their own (e.g. Kollar et al. 2006; Weinberger et al. 2007), guidance is needed to ensure positive learning outcomes. Designing and implementing didactic support for collaborative learning activities requires significant personal effort by teachers. Besides the preparation of content such as learning material, they also have to capture information on learners' knowledge, allowing them to form appropriate learning groups or to provide feedback to students about their knowledge.

Consequently, the aim of this article is to weigh the advantages and disadvantages of different measures to shape collaborative learning, reasonable efforts on the part of the teachers, and technical support possibilities in the context of educational classroom scenarios. Thus, we intend to combine the potentials of guidance (as a support for students) with largely automated technologies for identifying, transforming, and visualizing information on knowledge (as a support for teachers) into one tool that fits in regular school routines.

Starting from the theoretical background, we argue for the design of a tool that combines implicit and explicit guidance mechanisms relying on data that emerges from the application of largely automated text mining methods. Following this line, we experimentally evaluate the overall impact of the tool and its various features on students' learning in classrooms and then discuss the results.

Theoretical and conceptual background

Guiding learners explicitly and implicitly

Social interactions will not occur automatically just because they are possible or enabled in a given educational scenario (Kreijns et al. 2003). Frequently, guidance is needed to improve the quality of students' collaborative learning processes (Weinberger et al. 2007), such as structuring learners' interactions in specific ways. Since the extent to which students can be instructed varies considerably (Hesse 2007; Scardamalia and Bereiter 2014), teachers are free to choose from various options ranging from explicit to implicit guidance measures.

Explicit guidance means giving the learners a detailed specification of the collaborative process, e.g. in the form of scripts (cf. Fischer et al. 2013). Collaboration scripts are akin to movie scripts that prescribe in great detail how the actors have to perform (Dillenbourg 2002). Following Kobbe et al. (2007), such scripts consist of several components (activities, participants, roles, groups, and resources) and mechanisms (task distribution, sequencing, and group formation). Their core component is a learner's activity (Kobbe et al. 2007) that might be cognitive (e.g. explaining or asking), metacognitive (e.g. monitoring) or social (e.g. playing different roles) (Mäkitalo-Siegl and Kollar 2012). To foster the desired behavior of learners while they perform

their activities, teachers can particularly utilize the mechanism of group formation that determines how participants are distributed across the groups (Kobbe et al. 2007).

Common and frequent approaches to group formation target a heterogeneous distribution of participant characteristics in the groups. Dillenbourg and Jermann (2007) distinguish between two variants: (a) forming pairs or groups of learners with complementary knowledge, or by providing teammates with complementary information, and (b) forming groups based on learners' conflicting opinions. The members of complementary groups are expected to compensate for the gaps in different knowledge by exchanging and explaining missing concepts to each other. Examples of providing complementary information are the Jigsaw script (Aronson et al. 1978) and the Concept Grid (Dillenbourg and Jermann 2007). The UniverSanté script (Berger et al. 2001) is another example in which the complementary knowledge distribution is not induced (as in the jigsaw), but natural differences in knowledge are exploited. Here, collaborators studying in different countries contribute their knowledge about their national health system.

Based on complementary knowledge, learners may form different or even conflicting opinions. An example: A student who knows about the environmental effects of coal-based generation of energy may be inclined to accept nuclear power plants with all their risks. Another student with knowledge about wind energy and its advantages in comparison to nuclear energy will probably be more skeptical towards nuclear energy plants. Such heterogeneous opinions or incompatible conceptual views, though possibly leading to controversies (Johnson and Johnson 1979), are also promising for fostering collaborative learning, provided that the learners' abilities are homogeneous (O'Donnell and O'Kelly 1994). It is theorized from a developmental perspective (Piaget 1950; Flavell and Botkin 1968; Kohlberg 1969), from a cognitive perspective (Berlyne 1966; Doise and Mugny 1984), and from a social perspective (Johnson 1970; Johnson and Johnson 1979; Johnson 1980) that constructive controversies between learners can cause intellectual conflicts which in turn trigger positive learning outcomes (cf. Johnson and Johnson 2009). Here, structured controversy (Johnson and Johnson 1979) is an adequate example for the utilization of such grouping mechanisms, since it uses different (predefined) opinions as a basis for discussion. Another example is the ArgueGraph script (Dillenbourg and Jermann 2007) that is based on actually different opinions collected by means of a questionnaire.

While explicit guidance strongly shapes the way in which learners interact with each other, collaborative learning processes can also be implicitly guided by suggesting certain ways of thinking, communicating, and behaving, without directly instructing learners to perform specific activities (Bodemer 2011). Following Hesse (2007), the advantage of implicit guidance that initiates activities instead of commanding them lies in the learners' opportunity for self-regulation, which can increase motivation (Ryan and Deci 2006).

Examples of implicit guidance measures that have been proven to affect learning outcomes positively are representational guidance (e.g. Suthers 2001; Suthers and Hundhausen 2003) and cognitive group awareness (e.g. Bodemer 2011; Sangin et al. 2011). Representational guidance is based on the idea of making some parts of shared knowledge more salient than others so that they are more likely to be topics of discussion (Suthers 2001). Similarly, cognitive group awareness describes the awareness of relevant information that learners possess about their group members' state of knowledge, interests, assumptions or opinions (Janssen and Bodemer 2013). Accordingly, cognitive group awareness tools provide such information on learning partners (Bodemer and Dehler 2011; Janssen and Bodemer 2013), since they transform socio-cognitive variables in a way that can be fed back to the group (Bodemer and Buder 2006). Providing learners with specific information related to the

knowledge of other learners can reduce unnecessary workload and foster engagement in activities of collaborative elaboration in at least three ways (cf. Dillenbourg and Betrancourt 2006; Bodemer and Scholvien 2014). First, as such information refers to specific content, it may list and cue essential information, and thus can help to organize each learner's knowledge and focus the learning partners' communication. Second, being aware of other learners' knowledge can facilitate important grounding and partner modeling processes on the level of small or large groups. Third, when information is provided in a way that allows for comparing the learning partners' knowledge, learners can be triggered to discuss interdependent knowledge or controversial opinions, which can be particularly beneficial for learning as argued above for explicit guidance support. In the case of group formation based on knowledge differences, the additional support of cognitive group awareness thus seems reasonable. Tools supporting cognitive group awareness do not only assist students in discovering and closing gaps in their knowledge (e.g. Sangin et al. 2011), but also can lead to collaborative elaboration processes (Buder and Bodemer 2008), in which knowing about shared and unshared knowledge resources can trigger discussions about topics, with which only one learner in a group is familiar (Schittekatte and Hiel 1996).

Summing up, there are two promising ways to support collaborative learning on the basis of information on learners' knowledge: explicit, by forming knowledge-heterogeneous groups, and implicit, by providing learners with information on other learners' knowledge (including information cueing, cognitive partner modelling, and opportunities for comparison). From a teacher's perspective, applying such support in a classroom implies two challenges in particular:

(1) *Capturing and transforming information on learners' knowledge* (what and how much do learners know?). Group formation is an explicit guidance mechanism challenging teachers not only to choose the appropriate formation out of the numerous possibilities, but also to exert a great effort, if they do not randomly or artificially distribute the participants across groups. Instead, they have to gather information on what and/or how much learners know and to transform this information into difference values between students as a basis for heterogeneous group formation. The same applies to implicit guidance, since the creation of graphical representations requires appropriate input. In existing support measures that have proved successful in enhancing collaborative performance, categories for indicating learners' knowledge follow a top-down approach (cf. Dehler et al. 2011). This means that teachers have to determine beforehand what aspects of subjects to be learned are relevant, risking to miss other knowledge of their students on the subject area. A bottom-up approach based on student-generated content can reduce such problems. Furthermore, during collaboration, it might help students to connect the discussed topics to their own prior knowledge and motivate them by rediscovering self-generated instead of teacher-generated structures. Such structures based on students' knowledge have been gathered from different sources: based on students' subjective evaluation, for example self-assessment of knowledge (e.g. Dehler et al. 2011), or on objective indicators, e.g. results of a knowledge test (e.g. Sangin et al. 2011). Either way, the gathering and transformation of said information might contain errors caused by self-estimation or be burdensome, e.g. if teachers have to design and implement specific knowledge tests. An alternative approach would be the usage of student-generated text as a source for students' assessment, since the production of text is common in school and provides a data basis for being objectively evaluated.

(2) *Representing learners' knowledge.* After the transformation of knowledge, the teachers need to instruct the learners on how to collaborate. Implicit guidance can assist here with graphical representations, but challenges the teachers to visualize previously generated cognitive variables in a way that allows for easy information extraction and comparison. Indeed, they can choose common types of presentations that meet both of the aforementioned demands, but even the visualization of cognitive information using bar graphs is still time-consuming for teachers.

In conclusion, a combination of heterogeneous group formation and support of cognitive group awareness in school routines would be advisable for improving students' collaborative learning, but it comes with burdens for the teachers because it requires them to capture, transform, and visualize information about learners' knowledge. This raises the question of how to support the teachers. Taken all together, the answer could lie in automated technologies that allow for transforming given educational data, such as written homework or essays, to cognitive variables that can be used to form groups of learners with heterogeneous knowledge (or opinions), and for visualizing information on what and how much they know.

Text mining as a basis for forming groups and representing cognitive information

In the course of their school careers, students produce a huge amount of written homework such as essays or wiki entries. These texts are externalizations of knowledge and provide unstructured data to be analyzed. Text mining is based on the idea that cognitive information can be extracted from text, or rather be analyzed through largely automated techniques that turn text into 'numbers' (Miner et al. 2012). This means that unstructured textual data – words, sentences, paragraphs or documents – will be transformed into quantitative data representing cognitive information by values. In an educational context, examples for searching or sorting documents are the application of information retrieval in tutoring systems, e.g. Glosser (Villalon et al. 2008). The AutoTutor system (Graesser et al. 2004) uses Latent Sematic Analysis to analyze learner utterances with a natural language dialogue interface. Rosé et al. (2008) have applied automatic text classification techniques to CSCL corpora as a promising alternative to human coding. Other approaches have used document sorting techniques for clustering e-learning resources according to their similarity (e.g. Tane et al. 2004; Hung 2012). The analysis of words has recently been used to support teachers in monitoring their students, e.g., for identifying misconceptions from comments annotated to learning videos (Daems et al. 2014) or for analyzing conceptual change over time (Sherin 2012; Southavilay et al. 2013). Although not established in the context of cognitive group awareness so far, text mining appears to be very suitable for transforming students' written text into cognitive variables as a basis for forming groups and for visualizing cognitive information.

To promote learning processes it thus appears reasonable to equip a tool supporting cognitive group awareness with different automated functions focusing on clustering. Group formation is ideally based on document clustering, since each text or document represents an author or learner. Accordingly, we receive values that represent the similarity (or dissimilarity) between learning partners. Graphical representations are related to individual cognitive information and should therefore be based on concept extraction. Words will be clustered instead of documents, resulting in values that can be visualized as bars to best serve the instructional purposes of comparing components (Lee and Nelson 2004) and representing how intensively a

learner discusses an issue within his or her text. Depending on the method selection, such clustering procedures result in disjoint or non-disjoint clusters. This differentiation is of particular relevance regarding human interpretability of clusters: Disjoint results might provide only a small number of distinct clusters that are easy to interpret for that reason, whereas non-disjoint results might identify more existing clusters but also give more room for (possibly inaccurate) interpretation. In the following, we describe different text mining methods such as those based on Vector Space Models (VSMs) or Probabilistic Topic Models (PTMs) to discuss their suitability for the purposes described above.

VSMs treat texts as vectors in an n-dimensional vector space (Rajman and Vesely 2004) that can be represented using a document-term-matrix. The cells of such a matrix contain either the occurrences of words in a document (for instance, their frequencies), or their weighted values in a document (Miller 2005) on which statistical calculations can be performed. Therefore, VSMs have been combined with cluster analysis (e.g. Sherin 2012), factor analysis (e.g. Leydesdorff and Hellsten 2006), or singular value decomposition (e.g. Deerwester et al. 1990). These analytical methods allow for clustering words or documents (AlSumait et al. 2010). The latter technique has been successfully used for forming groups, e.g., by Manske et al. (2015) who used VSM based Euclidian distance, a proximity measure of cluster analysis, for group formation. Concerning word clusters, Sherin (2012) demonstrated that a simple VSM – extended by a cluster analysis and applied to segmented transcripts of students' explanations about the changing of the seasons – can identify learners' concepts and their dynamics over time. Certainly, other techniques for generating clusters without (hierarchical) overlaps should be considered. For example, factor analysis is a candidate, since Sherin's (2012) approach of combining the VSM with hierarchical centroid clustering only yielded acceptable results if it was additionally combined with a calculation of deviation vectors.

The currently most used instance of PTMs is Latent Dirichlet Allocation or LDA (Blei 2012). The idea behind this is that texts are represented by a distribution of different latent topics, where each topic is represented by the distribution of words in these texts (Blei et al. 2003). Thus, using LDA allows for identifying which parts of a text represent specific topics in a relative view. Such topics can be used as a basis on which distance measures can be calculated. Further, Southavilay et al. (2013) showed that LDA applied to collaborative writing data in GoogleDocs was suitable for creating topic evolution charts that clearly depict how topics evolve over time. Besides allowing teachers to monitor their students with charts, it is also a future objective to use this technique to make students aware of their development.

So far, the potential of text mining in educational contexts has not been fully exploited. The given examples show that text mining methods are not only suitable for grouping learners, and transforming qualitative data into quantitative data; they also allow for visualizing topic distributions and the evolution of topics over time. On this basis, using text mining to support cognitive group awareness can provide two important elements: (1) cognitive information about students, which is externally represented as a source of "content awareness" (what knowledge?) and "context awareness" (how much knowledge?), and (2) the heterogeneity of students' texts as a basis for group formation. As for (1), LDA and VSM extended by further analysis are candidate techniques that still need to be compared, since the latter, e.g. in combination with factor analysis, can provide disjoint topics whereas LDA yields topics interpreted on the basis of possibly overlapping top ranked words per cluster. Thus, the designation of resulting clusters and the distribution over topics might be different. On the one hand, a clear separation given through disjoint clusters can make an interpretation easier and provide precise values per topic. On the other hand, overlaps between multi-word topics

give additional means for interpretation and visualization. As for (2), the difference between LDA and VSM based analysis is that LDA generated data allow for determining a topic-wise distance between learners, whereas a VSM based calculation, e.g. Euclidian distance, can provide values based on distances of words. A word-wise analysis is more distinctive in terms of a sharper separation, but it is also conceivable that it may introduce artefacts due to the use of synonyms or filler words.

Overall, we assume that students profit from guidance while learning collaboratively. Regarding explicit or implicit guidance, we emphasized that the two forms of guidance provide different benefits for the support of students. Our approach is to combine grouping as an explicit guidance mechanism with the visual presentation of cognitive information to improve students' cognitive group awareness. However, from teachers' perspectives, the effort might appear to be unreasonable, since it is time-consuming and burdensome for them to gather quantitative information on learners' knowledge. As we have argued, text mining offers the advantage of automated extraction of cognitive information from student texts (e.g. from written homework). However, the potential of text mining is not yet sufficiently explored in the context of structuring collaborative learning. In the light of these considerations, we state the following research questions:

Research question 1: *Comparing and evaluating LDA and VSM based analysis – Which is more suitable for transforming and visualizing cognitive information (RQ 1.1) and characterizing heterogeneity between learners for group formation (RQ 1.2)?*

Research question 2: *What is the impact on learning outcomes of a tool supporting cognitive group awareness with integrated text mining methods to generate a data basis for visualizing cognitive information and grouping learners?*

We investigate research question 1 in the section "Design of the Grouping and Representing Tool (GRT)", where we select the text mining methods used in the tool and explain its functions in more detail. The subsequent section, "Experimental study", serves to answer research question 2 aiming to clarify that there is an effect utilizing its entire set of features. Finally, we interpret the results and discuss the tool's effects on students' learning.

Design of the Grouping and Representing Tool (GRT)

Our approach combines the advantages of educational guidance measures with the potential of text mining into one tool that we call the *Grouping and Representing Tool* (GRT). The GRT is designed as a tool supporting cognitive group awareness and is equipped with text mining methods that can be applied to learner-generated content such as essays, reports, or wiki entries to generate content- and context-related cognitive information on learners. Based on this information, the GRT generates graphical representations and forms learning groups. The graphical representations visualize cognitive information in terms of topics (human interpreted and named clusters of terms and phrases) and topics' extent for each student. The list of topics allows for monitoring one's own topics against the whole group's topics. The topics' extent allows for comparing one's own state of information on a topic to a partner's state of information on the topic. The group formation is based on the heterogeneity of learners that is operationalized here as the topical dissimilarity of two persons' texts. In this case, not the words within the texts but the texts are analyzed for their similarity. Therefore, the application of text mining methods should result in two types of values: topic values that serve to visualize

students' cognitive information and dissimilarity values that are used to match heterogeneous learning partners.

Selection of text mining methods

There are many different text mining methods, but none have been used so far to determine cognitive group awareness. Therefore, in order to investigate which of them is the most suitable, we address the first research question, whether VSM based clustering or (tailored) LDA achieves the better results in capturing cognitive information (RQ 1.1) and differences (RQ 1.2). As a data basis for validating and selecting the appropriate text mining methods, we prepared five essays about global warming that cover 11 different topics in total. This number of essays and topics was sufficient to establish various differences between each pair of texts: topic-related differences (reaching from no complementarity to a very high topical complementarity) and differences of opinion (given or not). Regarding topic-related differences, we calculated complementarity values (number of unequal topic occurrences per pair of texts divided by the total number of topics) as a standard to compare to. Regarding opinions, we could differentiate between pro and contra attitudes about stopping the causes of global warming. We systematically evaluated the results of text mining after applying (a) a VSM-based factor analysis in comparison to a LDA combined with a Gibbs sampling for capturing cognitive information (topic values) and (b) a VSM combined with Euclidian distance and LDA tailored by a topics' extent measure for capturing differences.

The quality of transforming written information into quantitative data and visualizing it (RQ 1.1) is linked to the precision with which a text mining method can identify topics and the topics' extent within a prepared text corpus. The results show that both text mining methods are suitable for identifying essential aspects of the given essays, provided that the paragraphs of the essays are segmented into single texts before clustering. Under this precondition, an explorative factor analysis with scree-test captured 8 out of the 11 identified subtopics. LDA delivered the same result, but was also able to capture topics that factor analysis could not identify, if the number of clusters had been manually increased to 11. We have to take into account that multiple runs are needed to determine the proper number of clusters and that the number of iterations must be high enough to ensure the internal consistency of the variable. For visualizing the extents of topics (which are equivalent to the quantitative values of clusters), LDA also performed best in a systematic comparison because it could reproduce the topics' extent in 4 essays adequately to human estimate, and also map the close similarity of two nearly identical texts better than factor analysis. Only one text could not be represented in an adequate way, since it contained a topic that appeared only in this text and in none of the others. Bearing in mind that the topical scope should be determined, we have chosen LDA for transforming and visualizing cognitive information with the tool.

Capturing differences accurately (RQ 1.2) is a precondition for successfully forming groups of discussants with heterogeneous knowledge or opinions. As can be seen from Table 1, a comparison of both text mining methods showed that the complementarity between two essays is better identified in the case of using a VSM combined with a calculation of Euclidian distances than by applying LDA and additionally subtracting the topics' extent of one text from that in another text. Only in one case VSM based values differed from the given order of complementarity ranks, whereas LDA based distances differed in several cases from the ranks and could not determine the extrema correctly. However, due to missing indicators for clearly identifying opinions from given clusters we did not pursue a group formation based on the

Table 1 Distance values between texts representing topical differences

|	text 2			text 3			text 4			text 5		
	$d_{Eucl.}$	d_{LDA}	c	$d_{Eucl.}$	d_{LDA}	c	$d_{Eucl.}$	d_{LDA}	c	$d_{Eucl.}$	d_{LDA}	c
text 1	**21.5**	**0.2**	**0.0**	31.0	1.3	0.6	33.0	1.3	0.7	**33.5**	1.4	**0.9**
text 2				31.3	1.4	0.6	33.1	1.3	0.7	33.2	1.2	**0.9**
text 3							27.0	0.9	0.1	29.6	**1.6**	0.3
text 4										23.9	0.9	0.2

$d_{Eucl.}$ = Euclidian distance. d_{LDA} = LDA based distance. c = complementarity value. The higher the value, the greater is the distance or complementarity. Lowest and highest values for each measurement method are bold

identification of controversial opinions. Taking into account all these results, we have chosen the approach of combining VSM with Euclidean distance to find heterogeneous and preferably complementary pairs, not excluding that heterogeneous text might also foster different opinions.

Specification of the functions of the GRT

To use the GRT in a classroom, we refer to the major tasks to support cognitive group awareness: transforming and visualizing cognitive information. Considering the key tasks and insights gained from the pre-investigation, we have designed the functions and sub-functions of the GRT as depicted in Fig. 1.

Function 1 serves to transform qualitative into quantitative data and consists of several text mining steps. Before applying text mining methods, we have to prepare the text corpus for subsequent analyses. Since the pre-investigation has shown that the identification of topics works much better if the paragraphs of a text are analyzed individually, all texts are segmented into smaller parts based on the count and length of paragraphs. To give an example: a document with eight paragraphs will be divided into eight single texts. Furthermore, we apply Natural Language Processing and Information Extraction methods. This means unneeded and unwanted items are removed, such as extra spaces, special characters, and punctuation marks. Also, we delete umlauts and replace affected letters by digraphs (e.g. ä will be replaced by ae), switch all letters to lowercase and reduce words to their root stem by removing suffixes based on the Porter Stemmer (Porter 1980). In addition, we generate an n-gram and a term list. The n-gram list identifies co-occurring words including their frequency of co-occurrence. We use this list as a basis to manually create a thesaurus that combines syntagmatically cohesive words to one phrase by combining the terms with underscores (e.g. *travel by bus and train* will be combined to travel_by_bus_and_train). Concurrently, paradigmatically synonymous terms and n-grams are identified manually and included in the thesaurus (e.g. the n-gram *use public transport* will be transferred to its synonym travel_by_bus_and_train). The term list contains all words occurring in the text corpus, including their frequency of occurrence and tf/idf values, which are calculated by dividing the inverse document frequency from the term frequency. It is used to detect single words with paradigmatically synonymous relation (e.g. *automobile, auto, motor car* and *car*) and to add them to the thesaurus to be manually replaced by the more standard form (in this example: *car*). After applying the thesaurus and deleting stop words based on the stop word list of the *R* text mining package *tm 0.5–8.3* (Feinerer and Hornik 2013), we re-generate the term list again, including the frequency of occurrence and tf/idf values for each term and phrase. Following Leydesdorff and Hellsten
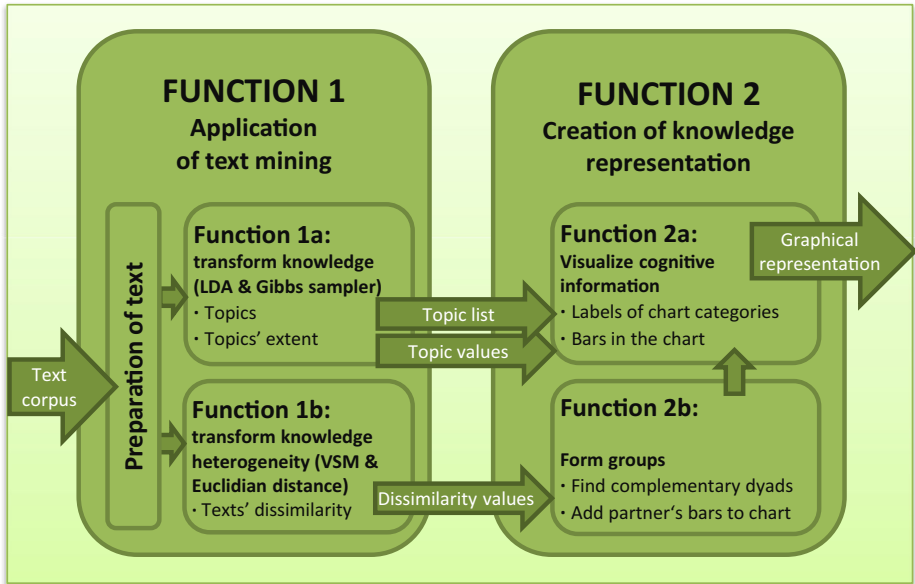
Fig. 1 Schematic diagram illustrating the functions of the GRT: Function 1 embeds text mining which is needed for extracting concepts as well as transforming cognitive information from text into concept clusters (topics) and topics' values (1a), and transforming heterogeneity in dissimilarity values (1b). In function 2, the GRT visualizes charts with identified topics as labels and topic values as bars (2a), and forms heterogeneous dyads from the dissimilarity values (2b). The result is a jointly graphical representation for each heterogeneous dyad

(2006), who suggest determining the relevance of terms according to their frequency within a text corpus, we remove all words with an overall frequency and tf/idf value less than a self-determined value dependent on the characteristics of the whole text corpus. Prepared texts resulting from this procedure serve as data basis for executing function 1a and 1b.

To transform cognitive information (function 1a) from the segmented prepared text, we use LDA combined with Gibbs sampling as it is included in *R's* lda package. The lda package allows determining the following parameters: the suspected number of clusters K, the number of iterations, alpha – the scalar value for the topic distribution – and eta – the scalar value for topic multinomial. K depends on the number, length, and contents of texts in a corpus and needs to be varied in several passes to determine an appropriate number of topics. Following the pre-investigation, it proves useful to choose the average number of paragraphs within a text corpus as a minimum starting point. Then, we command to output the *top ten words* in a table that forms the basis for human interpretation of the clusters as topics by naming them. In addition, we save the text-topic-matrix with its cells containing the frequency of term and phrase assignments to each particular topic. These values represent the topic values. Since the texts per person are segmented, we sum the values thereafter for each student and each topic. Finally, there are as many values as indicators for each student's topics' extents available as previously determined with K. The list of named topics in combination with topic values represents the thematic focus of each text as an indicator of the student's knowledge. After an additional weighting of these values on the total number of terms and phrases in the given prepared text, they serve as the basis for visualizing the graphical representation (function 2a). For extracting and transforming complementarity (function 1b) from the prepared texts (not segmented this time), we use the built-in R function "Distance Matrix Computation" with the selection of the method "euclidean" that calculates the Euclidean distance between two vectors,

each vector representing a prepared text or person. The greater the value, the higher the distance between two vectors. These dissimilarity values serve as the data basis for deciding which learning partners to match into a dyad (function 2b).

The visualization of cognitive information on students (function 2a) is based on the results of function 1a. We use them to list the names of identified topics as text and to display each student's topic values as bars in a chart. Existing cognitive group awareness tools provide different levels regarding the comparability of cognitive variables. Bars (cf. Sangin et al. 2011) or diagrams visualizing awareness information topic-wise (cf. Dehler et al. 2011; Bodemer 2011) are used in tools aiming to facilitate comparisons between learners, while other tools provide complex representations such as concept maps self-generated by the students from learning text (e.g. Engelmann and Hesse 2010) that are difficult to compare. We chose bars for visualizing topics' extents, since they are common in given tools and recommended to be used for comparing at least two components (Lee and Nelson 2004). Figure 2 shows an exemplary bar chart for one person. This represents an intermediate state, since we will further add a learning partner's bars to the chart.

To group students into maximally heterogeneous learning dyads (function 2b), we scan the distance matrix for its highest value. We start matching the two persons who are authors of the respective texts (one text heading the column, the other heading the row that belongs to the cell concerned in the distance matrix) into a dyad, resulting in a jointly graphical representation based on the visualizations from function 2a as can be seen in Fig. 3. Once combined, the dyad's vectors will not be integrated in follow-up scans of the distance matrix for the next highest value in this iterated process. The feedback of the graphical representation is embedded in school classes, which is why it
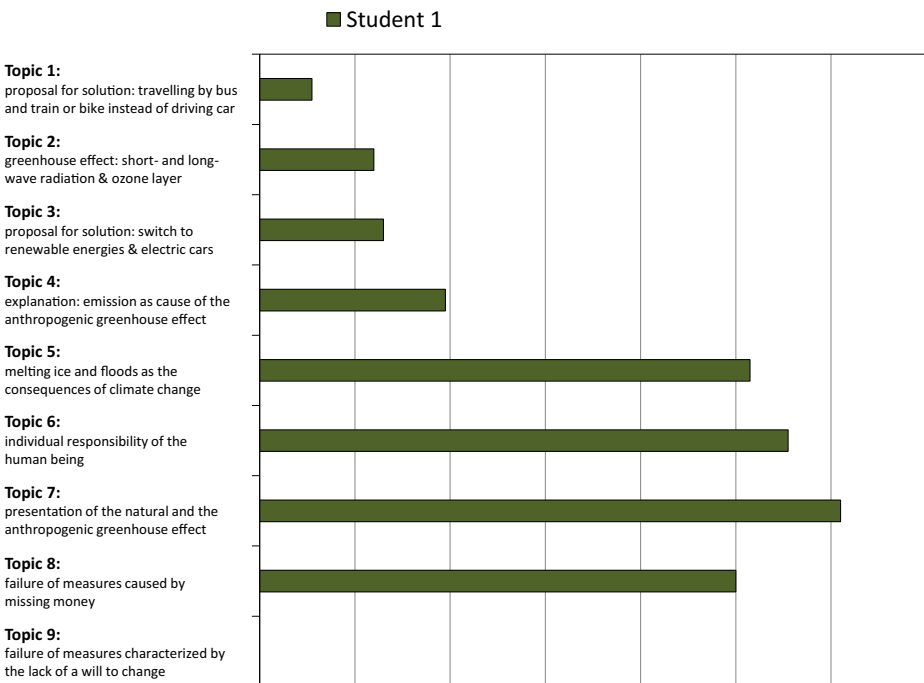
Fig. 2 Example of a graphical representation for one person generated with the aid of function 2a of the GRT. On the left is the list of identified topics generated employing function 1a. The learner's topics' extents are comprehensible on the basis of the bar lengths in the bar chart on the right

needs to be accompanied by further instructions depending on the classroom scenario. In either case, it is presented with the explanation that the graphical representation is based on students' own contents and gives an overview of the topics covered in all texts.

Experimental study

The experimental study refers to research question 2 and serves to explore how this type of tool can support learning. Therefore, we measure the effect the GRT might have through forming heterogeneous groups of students and providing them with graphical representations that allow for comparing their cognitive information. We have used the tool in an upper secondary school class. Figure 4 illustrates the main procedure (including the GRT as supportive element and as tool for data analysis, too), which we present now so that the explanations in the following sections are better understandable.

Research design

We investigated the effect of the tool in a real classroom setting asking students to write an essay on *global warming* and to discuss their essays' topics afterwards. We used a 2 × 2 mixed factorial design with randomly assigned group membership as between-subject factor (group supported by the GRT vs. group not supported by the GRT) and the phase as within-subject factor (writing phase vs. modifying phase). On the one hand, we tested the effect on learners'
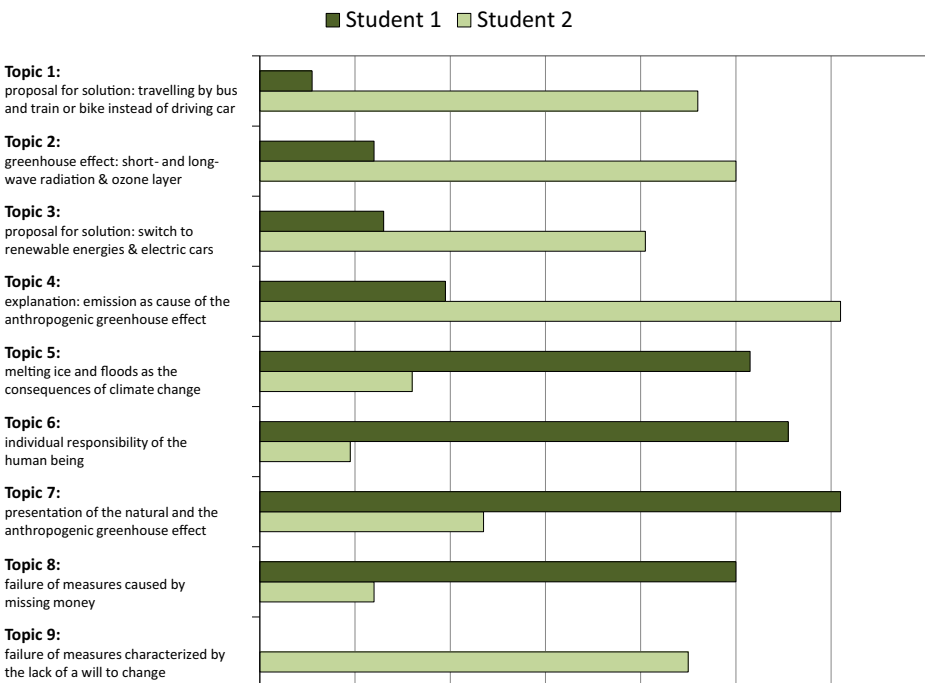
Fig. 3 Example of a graphical representation as it would be provided to the learners. It allows both learners to compare themselves to the whole group by recognizing the topics on the left. Further, they can compare each other's knowledge based on the bar lengths
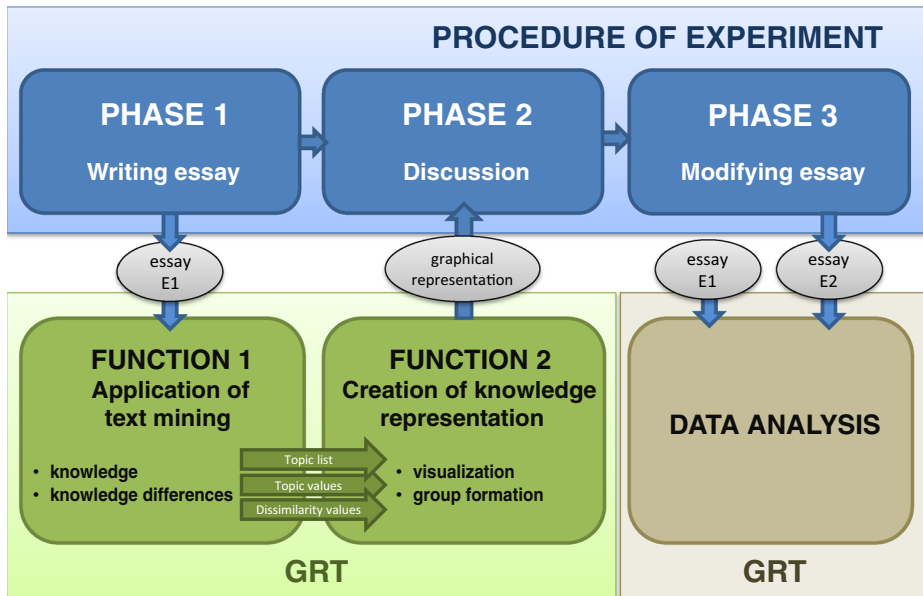
Fig. 4 Overview of the experiment's process of embedding the GRT. The above component depicts the structuring of the experiment, the component in the lower left part represents the GRT as it was used for creating the graphical representations, and the component in the lower right describes the data analysis process that is further explained in the method part

knowledge based on the optional modification of their essays after discussion; on the other hand, we observed the change of knowledge-related differences between learning partners, particularly their convergence to each other.

In addition, we controlled the effect of Euclidian distance as it is used for matching the dyads with the GRT on knowledge acquisition. Therefore, we used a between design with group membership as between-subject factor again, but the additional calculated knowledge acquisition value as dependent variable this time.

Participants

The sample consisted of 56 male (53,7 %) and female (46,3 %) high school students of a German upper secondary school ("gymnasium"), 15 to 17 years old ($M = 15.34$, $SD = 0.51$), and attending one of three classes with a main focus on geography. The students had been brought up to the following knowledge level about *global warming* at the start of the survey: they know the climate system and its components. Furthermore, they are capable of distinguishing between natural and anthropogenic greenhouse effect and can address the root causes of *global warming*. For the purpose of conducting the study, students were assigned to dyads as described in the section "Learning dyads resulting from GRT's grouping and from randomized grouping".

Dependent variables & hypotheses

Dependent variables are knowledge represented in text and heterogeneity. As can be seen from Fig. 5, respective values are generated with function 1 of the GRT again, but here both student
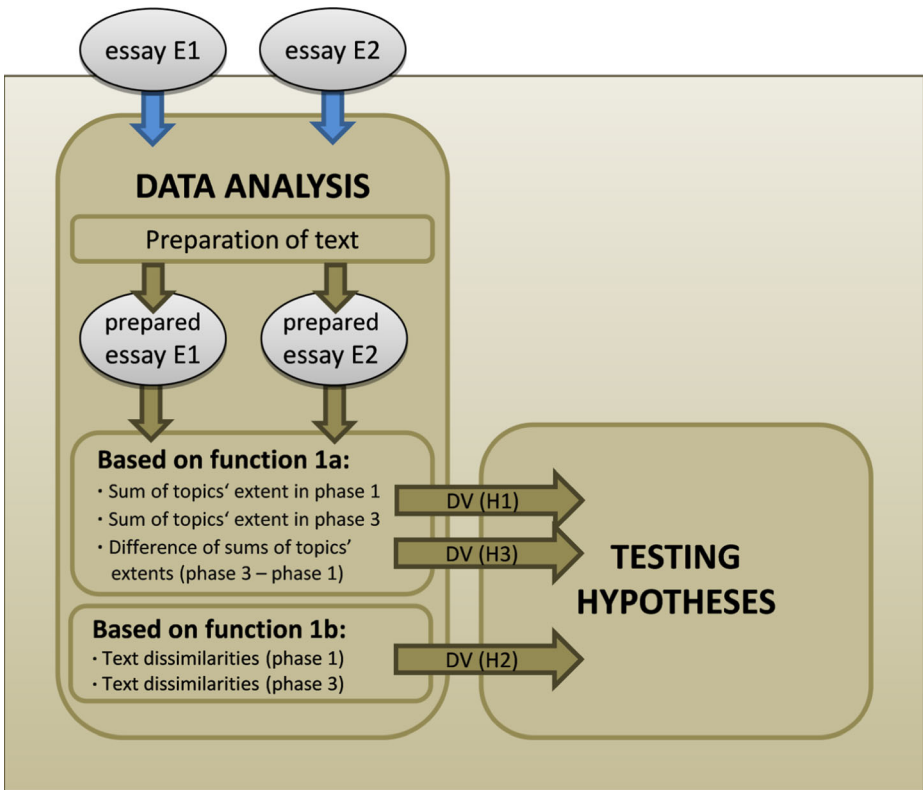
Fig. 5 Diagram depicting the GRT's use for data analysis. For calculating the dependent variables (DVs), we execute function 1 of the GRT again. As a result, we receive topic values and dissimilarity values as illustrated in Fig. 1. Based on the topic values resulting from function 1a, we generate the sum of topics' extent for phase 1 and 3 as well as the difference of the sums of topics' extents. Based on function 1b, we gain two dissimilarity values this time, one representing knowledge differences before discussion and another one representing differences after discussion

generated essays are the basis for the analysis – the initial text version and the modified text version.

We operationalize knowledge represented in text as the sum of topics' extent per student and per phase, but without weighting the topic values this time so that we are able to compare the topics' extent at different times. The sum of topics' extent is related to the number of words per text, but differs from it insofar that only relevant terms and phrases are assigned to different topics and thus are adopted in the sum of topic values. Heterogeneity is defined as the Euclidian distance between dyad members. To determine each dyad's dissimilarity value, we weight terms and phrases per essay on the length of the respective text to make the values comparable before calculating the Euclidian distance as described in the last section.

Based on this operationalization, we segmented research question 2 asking for the effect of the GRT into three main hypotheses investigating the learning effect (hypothesis 1), the convergence of group members (hypothesis 2), and the control for heterogeneity as confounding treatment factor (hypothesis 3). Since collaborative learning improves learning outcomes (Johnson et al. 1998; Springer et al. 1999), we first assume that students from both groups represent more knowledge in their text after discussing with their learning partner, expressed in

the fact that students allowed to modify their essay take this opportunity to add new text to their document. We attribute the expected increase of written knowledge to the impact of the respective collaboration partner, also assuming that teammates will converge after discussion. Furthermore, we expect both effects described above to be particularly strong when students are supported by the GRT. It is an explicit feature of the tool to assign students of high diversity to groups (based on Euclidian distance measures). Seen in the terms of implicit guidance, the GRT's graphical representations further visualize a list of topics (resulting from clusters of terms and phrases) supporting cognitive group awareness, since the list informs learners about the whole groups' topical scope and allows them for comparing the others' scope to their own essay's scope. Additionally, the representations contain bar graphs (resulting from calculations of each student's topic's extent) also supporting cognitive group awareness, since the bars provide knowledge-related information on the individual learning partner per topic and allow students for comparing their own to the teammate's topic's extent. Overall, we assume about the entire set of GRT's functions to have an impact on learning, since the visualization of different knowledge can trigger collaborative elaboration processes while discussing (Buder and Bodemer 2008) and elicit better learning results than homogeneous knowledge (Manske et al. 2015). Further we expect the GRT's functions to increase knowledge convergence, since the feedback of different states of knowledge usually leads to a discussion about unshared knowledge between learning partners (Schittekatte and Hiel 1996) and motivates the discussants to fill their knowledge gaps (Sangin et al. 2011). Derived from the above considerations, we hypothesize:

Hypothesis 1a: *While modifying the essay, students' text represents more knowledge than during the writing phase.*

Hypothesis 1b: *Students supported by the GRT learn more between writing and modification of the essay than students without support of the GRT.*

Hypothesis 2a: *While modifying the essay, dyads' heterogeneity is lower than during the writing of the essay.*

Hypothesis 2b: *Dyads supported by the GRT show a higher convergence between writing and modifying the essay than dyads without the support of the GRT.*

As can be seen from the explanations above, we intend that the effect of the GRT, on the one hand, is based on the heterogeneous group formation and, on the other hand, on the visualization of cognitive information. Referring to this confounding of both variables, we expect a positive effect of the groupings' dissimilarity values on learning:

Hypothesis 3: *The effect of the GRT on knowledge acquisition becomes larger, the greater the heterogeneity within dyads.*

Materials & procedure

The survey was realized in two computer rooms enabling a controlled conducting of the experiment. All students worked under the same conditions during their regular school hours

and were supervised by a teacher and two researchers, who were allocated to the two rooms. Each student had a single computer available on which the homepage of SoSci Survey, a software package for professionally conducting online surveys, was called up at the beginning of the lesson. The pages had been individualized in accordance with the experiment and contained all instructions and questionnaires for carrying out the study. Both computer rooms were occupied for three hours on two different days per class, which meant six on-site appointments in total. Phase 1 was performed in a single lesson (45 min); for phase 2 and 3, a double lesson was scheduled (90 min).

As a preparatory measure, teachers assigned homework in which the students had to reflect on the extent to which nature and human kind are triggers for *global warming*. What do they know from books and films on the subject? What emotions do they personally have? What are the arguments of climate skeptics? And what countermeasures against *global warming* consequently result from it?

Phase 1, the first part of the survey, was carried out one week after assigning the homework. After providing their personal data (full name, name of the teacher, sex, age), the students had 40 min to write an essay about *global warming*. Since it was a partial result of the pre-investigation that all texts might have the same structure, all students were given a brief guide about how to write a problem-solving essay. The following main topic should be focused on: "Global warming: what is the extent of its natural and man-made causes? And what counter-measures could be taken?" The text should be about 1.5 to 2 pages with thematic units distinguished by paragraphs. The text corpus consisting of all essays in phase 1 was the basis for the application of the GRT. We transformed the text into quantitative data as described in the section "Specification of the functions of the GRT" with the following parameters: in the text preparation step, we removed all terms and phrases with an overall frequency less than four and a tf/idf value less than 0.0011. For the application of Gibbs Sampling LDA, we selected nine clusters as a target parameter and iterated until we reached stable output. This was followed by the output of the *top ten words* per topic in a table for human interpretation. Written topic labels resulting from this human interpretation and an example of visualizing respective values can be seen from the graphical representations as shown in Fig. 2 (for a single student) and in Fig. 3 (for a heterogeneous composed dyad).

Phase 2 was performed together with phase 3 in a double lesson three weeks after phase 1 (after end of autumn holidays). Dyads were randomly assigned to experimental groups. The students found a printed version of their essays at their desk, which they were asked to read to recall its content. In addition, each dyad supported with the GRT had a graphical representation as described previously (see Fig. 3). This print included the explanation that the diagram is based on their individual essays and that it gives an overview of the topics that are covered in the whole text corpus. Furthermore, the dyads were informed that the bar length in their diagram is an indicator of how intensive they have addressed each topic and that they can use the bars to compare their own with their learning partner's cognitive information. Finally, the GRT group members received the written instruction to discuss their essays within the next 30 min with the given objective to learn from the partner as much as possible. This instruction was also given to the dyads in the unsupported condition, but without the diagram and explanatory text.

Phase 3 followed directly after the discussion. Now, each student had to work on the computer again; the digital version of the individual essay was presented to each student on the webpage. Additionally, we presented the written instruction to modify the text, if the

discussion with the partners resulted in new knowledge; modifying was more precisely defined as deleting, rewriting or completing its passages. Thirty min were provided for this purpose.

Results

Text corpus resulting from students writing and modification phase

The text corpus used for applying text mining methods consisted of 54 documents for each phase, which makes a total of 108 texts. After the segmentation of these texts following their division by paragraphs, 963 text fragments were available for analysis. Table 2 provides an overview of how many words were written per group (supported with the GRT and not supported with the GRT) and per phase (phase 1: writing phase before discussion, phase 3: modifying phase after discussion). Additionally, the respective range is specified. From the table it can be concluded that the number of words has increased on average after discussion. Nevertheless, it is worth noting that one text was shortened in the GRT-supported condition. In two cases, one in the GRT-supported condition and one in the other condition, the number of words did not change. The average number of paragraphs per text was nine.

After text modification in phase 3, the GRT's function 1 as it is depicted in Fig. 5 was applied to all essay fragments in order to capture the topical space of the whole text corpus. Table 3 illustrates resulting clusters with the most probable terms and phrases per topic, listed by their rankings. The newly generated topics conform to the topics captured in phase 1 to a great extent. In contrast to the nine clusters previously chosen for dyad matching and visualization of cognitive information, a number of 11 clusters was considered as appropriate for data analysis. Additional topics are: illustration of the carbon dioxide cycle (Topic 2) and international responsibility as a solution (Topic 5).

Learning dyads resulting from GRT's grouping and from randomized grouping

After applying GRT's grouping function in the supported group ($n = 26$) based on learners' texts from writing phase and assigning the learners in the control group ($n = 28$) to randomized groups, the investigation involved 27 dyads of balanced average age and ratio between female and male participants. The missing group in the GRT-supported condition (13 instead of 14 dyads) was because two students were

Table 2 Amount of written words per essay in phase 1 ($N = 54$) and phase 3 ($N = 54$)

phase	Supported group ($n = 26$)			Unsupported group ($n = 28$)			Overall ($N = 54$)		
	M	SD	Range	M	SD	Range	M	SD	Range
write (1)	357.58	84.42	200–521	359.14	103.74	176–529	358.39	94.05	176–529
modify (3)	460.31	128.19	227–841	412.14	120.67	221–592	435.33	125.53	221–841

M = mean. SD = standard deviation

Table 3 Identification of 11 topics within the essays about global warming in phase 1 ($N = 54$) and phase 3 ($N = 54$)

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
concepts	human_being	carbon_dioxide	global_warming	problem	germany
	problem	carbon_dioxide_emission	take_measure	current	use
	solution	tree	natural_and_anthropogenic _greenhouse_effect	purchase	china
	day	plant	problem	electric_car	consume
	help	airplane	cause	expensive	build
	car	burn	human	people	citizen
	try	human_being	have_dimension	pollutant	car
	mobile_phone	year	extent	car	country
	earth	power_plant	natural_greenhouse_effect	driving	nuclear_power_station
	establish	factory	greenhouse_gas	solution	renewable_energy
topic	Individual responsebility of human being	Illustration of the carbon dioxide cycle	Greenhouse effect as the reason for global warming	Electric car as (expensive) solution	International responsibility as a solution

Table 3 (continued)

	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10	Topic 11
concepts	global_warming	ride_bike	earth	earth	global_warming	wind
	human_being	bus_and_train	atmosphere	people	rising_sea_level	renewable_energy
	expensive	car	long_wave_radiation	change	water	coal_power_plant
	measure	drive_car	greenhouse_effect	many_people	country	energy
	mass_animal_husbandry	drive	short_wave_radiation	waiver	warming	solar_energy
	put_into_practice	trael_by_bus_and_train	radiate	human_being	climate	energy_source
	stop_global_warming	bike	anthropogenic_greenhouse_effect	think	earth	electricity_generation
	stop	electric_car	global_warming	give_up_things	flood	sun
	animal	walk	natural_greenhouse_effect	interest	pole_melting	build
	time	many_people	ozone_layer	winter	animal	reduce
topic	Failure of measures due to a lack of money	Proposal of solution: travel by bus and train or cycle instead of driving car	Greenhouse effect: short-wave and long-wave radiation and ozone layer	Waiver readiness as prerequisite	Melting ice, floods, and death of animals as a result	Proposal of solution: switch to renewable energy

Since the results are based on German texts, we have translated the concepts from German to English. The row 'concepts' illustrates which concepts LDA has assigned to which topic. The columns present the assignments of concepts to topics, sorted by their probability of occurrence in a topic in descending order. The row 'topic' explains and titles the topics that arise from the interpretation of concept clusters of the respective columns

reported sick in phases 2 and 3. As a result, the intended learning partners of the two missing students were spontaneously assembled into a new dyad and they consequently had to work with two different diagrams instead of a joint representation. Another disadvantage was that one of the initially intended dyads showed a maximum value in Euclidean distance. Its possible effect could thus not be taken into account. However, since we intended heterogeneity within groups to be part of the treatment, we had to ensure that the heterogeneity within dyads supported by the GRT still is higher than the heterogeneity within dyads in the unsupported group. A one-way ANOVA with group membership (supported by the GRT vs. unsupported by the GRT) as between-subject factor and the dissimilarity values, on which the group formation of heterogeneous dyads in the GRT is based, as dependent variable showed that the text dissimilarities within dyads of the GRT-supported group were not significantly higher than within dyads of the unsupported group, $F(1, 52) = 0.40$, $p = .531$, $\eta_p^2 = .008$.

Students' learning (hypothesis 1)

Starting with hypothesis 1, we intended to investigate student's increase of knowledge caused by the discussion (hypothesis 1 a) and the additional support by the GRT (hypothesis 1 b). Therefore, a two-factorial repeated-measures ANOVA was used with the phase (phase 1 vs. phase 3) as within-subject factor, group membership (supported by the GRT vs. not supported by the GRT) as between-subject factor, and sum of topics' extent as dependent variable.

Concerning the assumption that students present more knowledge in their text after discussion and thus have a higher sum of topics' extent in their modified essay than in their initial essay (hypothesis 1 a), there was a significant effect of the phase on students' sum of topics' extent, $F(1, 52) = 67.5$, $p < .001$, $\eta_p^2 = .565$. Table 4 illustrates that sums of topics' extent per text in phase 1 were lower than in phase 3, which means that there was a positive effect on learning caused by the discussion.

With regard to the expected interaction between the phase and the use of the GRT, it was assumed that especially students supported by the GRT add more relevant terms and phrases to their essays after discussion than students in the unsupported group do (hypothesis 1 b). There was a significant interaction effect between phase and group membership on students' sum of topics' extent, $F(1, 52) = 7.34$, $p = .009$, $\eta_p^2 = .124$. As can be seen from Table 4, participants supported by the GRT added twice as many relevant terms and phrases to their essays after discussing (30 % increase of knowledge represented in the text) than students in the unsupported group (15 % increase of knowledge represented in the text).

Table 4 Sum of topics' extent per essay in phase 1 ($N = 54$) and phase 3 ($N = 54$)

phase	Supported group ($n = 26$)		Unsupported group ($n = 28$)		Overall ($N = 54$)	
	M	SD	M	SD	M	SD
write (1)	97.5	24.7	98.3	26.7	97.8	25.5
modify (3)	127.1	40.0	113.2	31.7	119.8	36.3

M = mean. SD = standard deviation

Students' convergence of knowledge represented in the text (hypothesis 2)

Hypothesis 2 serves to investigate students' expected convergence. Therefore, the decrease of Euclidian distance caused by the discussion (hypothesis 2a) and the additional support by the GRT (hypothesis 2b) was observed. Again, a two-factorial repeated-measures ANOVA was used with the phase as within-subject factor (phase 1 vs. phase 3) and group membership as between-subject factor (supported by the GRT vs. unsupported by the GRT). As independent variable, convergence is linked to text dissimilarities (cf. Fig. 5).

Concerning the assumption that students converge to their respective learning partner during collaboration, it was tested, if dyads' Euclidian distance is higher before than after discussion (hypothesis 2a). There was a significant effect of the phase on the dyads' text dissimilarities, $F(1, 52) = 118$, $p < .001$, $\eta_p^2 = .694$. As shown in Table 5, average text dissimilarity between two discussion partners in phase 1 is higher than in phase 3, which means that dyads converged.

Regarding the expected interaction between phase and the use of the GRT, it was assumed that the convergence of dyads in the GRT-supported group is higher than in the unsupported group (hypothesis 2b). There was a significant interaction effect between phase and group membership on text dissimilarity between discussion partners, $F(1, 52) = 15.3$, $p < .001$, $\eta_p^2 = .277$. As can be seen from Table 5, participants supported by the GRT converged more from phase 1 to phase 3 (18 % decrease of text dissimilarities) than unsupported group members (6 % decrease of text dissimilarities).

The Euclidian distance measure is linked to dyads' convergence, but refers to the whole text with no distinction between different topics and thus without allowing to keep track of topics' development. To get greater insight into how the GRT influences the discussion partners' convergence, the dyad with the highest decrease of heterogeneity (from 18.62 in phase 1 to 12.84 in phase 3) was selected to visualize and learn from their change of topics' extent. Figure 6 shows a merging of their graphical representations that is based on the results from the analysis of the whole text corpus, including data from writing phase and modification phase.

In this example, we can see three different states of knowledge distribution in phase 1: (1) both learning partners have written about a topic (two plain bars per topic), (2) one has written about a topic, the other has not (one plain bar per topic), and (3) no one has written about a topic (no plain bar per topic). Concerning students' topical change over time, we can firstly state that in nearly all cases students' topics' extents have risen: given a total of 11 topics, values of both increased in five cases (both have bars with dashed lines per topic, cf. Topic 1, 2, 4, 5, 10) and values of one increased in four cases (just one has a bar with dashed lines per topic, cf. Topic 7, 8, 9, 11). Just in one case was there no change (cf. Topic 3). In another, we

Table 5 Euclidian distance between dyad's essays in phase 1 ($N = 54$) and 3 ($N = 54$)

phase	Supported group ($n = 26$)		Unsupported group ($n = 28$)		Overall ($N = 54$)	
	M	*SD*	*M*	*SD*	*M*	*SD*
write (1)	16.75	2.43	15.89	1.85	16.31	2.17
modify (3)	14.21	1.98	14.70	2.06	14.46	2.02

M = mean. *SD* = standard deviation. Values are multiplied with 100 for more comprehensible display
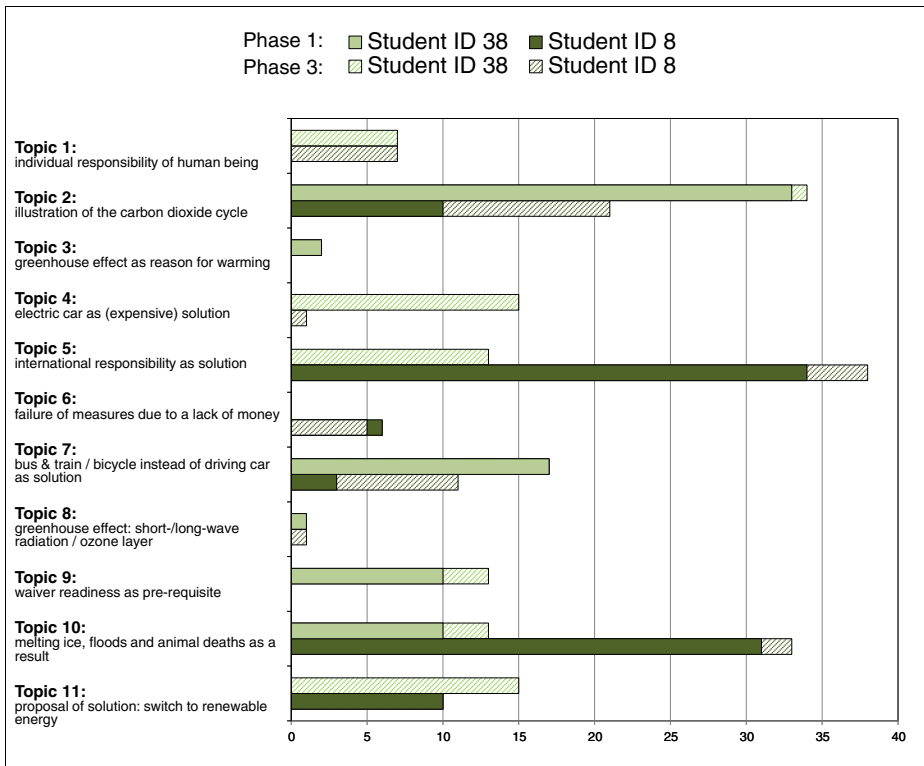
Fig. 6 Diagram merging graphical representation from phase 1 and phase 3, belonging to the dyad with the highest decrease of text dissimilarities between both phases. Plain bars illustrate topics' extents in phase 1, bars with dashed lines depict students' modifications, or better, their additions per topic within their essays

could even find a small decrease (cf. Topic 6). In the event of the initial knowledge distribution as described in (1) and (2), there mostly appears to be some convergence (e.g. Topic 2, 7, 8). In the case of a distribution as given in (3), both partners added text about belonging topics (Topic 1, 4). Besides the assumed guidance effect through comparing with the learning partner (comparison of bar lengths) and the whole group (recognizing topics without plain bars), it can further be noticed that topics based on opinions (e.g. Topic 1) seem generally to be more often discussed than topics based on domain knowledge (e.g. Topic 3).

Furthermore, the investigation of qualitative data reveals two main types of text modifications: the modification of text through (1) adapting the text to the topics of the whole group, and (2) adding concepts that were not part of the visualization but communicated by the learning partner. Student 8 shows type (1) modifications: he / she mainly added concepts that were visible in the topic list, e.g. taking Topic 1 into account, Student 8 added the sentence "Every human being can contribute to the fight against climate change, e.g. they can segregate waste, travel by bus and train or use power from renewable energy resources." [this and further examples are translated to English by the authors]. In doing so, the concepts of the learning partner were only partially taken into account, since the supplemented sentences were not always consistent with the learning partner's descriptions. For example, the sentence by Student 8 "Due to many emissions, the carbon dioxide gets caught at the ozone layer; solar radiation can no longer escape from the earth's atmosphere due to the carbon dioxide, causing

the Earth's surface to warm up." is not exactly the same as the description of Student 38 who already wrote about that topic in phase 1: "The climate increases because there is an ozone hole through which UV rays shine. Earth is warming, if they get into contact with the CO_2 released by man.". In this case, examples suggest that the topic list supports students to focus on topics for their written report or for the discussion with the learning partner.

Student 38 shows type (2) assimilations: he / she mainly added concepts that indeed were related to the topic list but not a part of it. Instead, we could find concepts in the modified text that were also part of Student 8's text in phase 1, e.g. Student 8 wrote about brown coal and natural gas as limited resources and recommended that they should be replaced by solar, wind and hydro power. Thus, Student 38 added the sentences "Nowadays, people use lignite and natural gas, but both are limited energy resources. Thus, renewable energy should be used instead (solar, wind and hydro power)." which are related to Topic 11 and further associated with concepts from partner's text. Another example of Student 38 is the addition of the concepts bioethanol and canola fields which were used by Student 8 in phase 1: "One could in general build more cars that run on bioethanol. However, one would have to grow many canola fields for this...". It can be seen here that the topic was refined and related to financial aspects by the author, since he / she further added "... and if the demand for these cars is too high, you would also need more workers to grow the canola fields and to produce the cars faster.". Student 38 wrote more about Topic 11 after the discussion than the learning partner did. Another proof of elaboration is the addition of concepts belonging to Topic 4. Here, Student 38 added a concept, electric car, that was named in the topic list, but further related it to other topics after the collaboration. "Instead of normal cars, one could also use electric cars, which are powered by batteries. However, the electricity to run these cars is probably not of renewable energy, what in turn is harmful to the environment.". Thus, it seems that the discussion induced Student 38 to relate electric cars with Topic 11 which is about renewable energy. Further, Student 38 discusses if the solution is too expensive and relates it to the emission of CO_2 (which is related to topic 2): "One could also buy a car that produces less CO_2, if electric cars are too expensive."

Overall, we can assume the following: Whereas type (1) modifications seem to result from comparison to the whole group (topic list), type (2) modifications appear to be an assimilation of new knowledge originating from the discussion with the learning partner.

Impact of group formation on knowledge acquisition (hypothesis 3)

Hypothesis 3 serves to investigate further the effect of the group formation on knowledge acquisition. Although there is no difference in dyads' text dissimilarities (representing dyads' heterogeneity) between experimental and control group (cf. section "Learning dyads resulting from GRT's grouping and from randomized grouping"), we assume at least that the effect of the GRT on knowledge acquisition becomes larger the greater the heterogeneity of a dyad is (hypothesis 3). The hypothesis was tested with a moderation model using SPSS. Group membership served as independent variable that was centered by contrast coding (not supported by the GRT = −0.5; supported by the GRT = 0.5). Knowledge acquisition (sum of topics' extent in phase 1 subtracted from sum of topics' extent in phase 3, cf. Fig. 5) was used as dependent variable and mean-centered text dissimilarity was examined as moderator variable. One statistical outlier was deleted to fulfill the assumption of normally distributed errors, but the analysis is also reported including said case.

The moderation model including both main effect terms and the interaction term was rejected (R^2 = .14, $F(3, 49)$ = 2.68, p = .057). Without interaction term, the model explained 14 % of the variance of knowledge acquisition caused by the collaboration (R^2 = .14, $F(2, 50)$ = 4.11, p = .022). The regression coefficient of group membership was estimated at 11.39 (SE = 4.66) and group membership significantly predicted knowledge acquisition (β = .321, $t(52)$ = 2.44, p = .018). The regression coefficient of dissimilarity was estimated at 1.53 (SE = 1.12) and dissimilarity did not significantly predict knowledge acquisition (β = .180, $t(52)$ = 1.37, p = .176).

Including the statistical outlier in the analysis, the moderation model explained 21 % of the variance of knowledge acquisition caused by the collaboration (R^2 = .21, $F(3, 50)$ = 4.53, p = .007). However, there was no significant effect of the interaction term. Again, the model without interaction term was chosen. Including only the main effect terms, the model explained 20 % of the variance of knowledge acquisition caused by the collaboration (R^2 = .20, $F(2, 51)$ = 6.40, p = .003) with both group membership (β = .327, $t(53)$ = 2.61, p = .012) and dissimilarity (β = .279, $t(53)$ = 2.21, p = .031) significantly predicting knowledge acquisition.

Discussion

Classroom orchestration comes with the challenge for teachers to find a balance between best learner support, and reasonable expenses, especially if teachers want to establish collaborative learning in school routines. To guide their students in order to ensure good learning outcomes, they have to take measures such as explicit group formation or the provision of visual representations as an implicit trigger for a richer elaboration. We have argued that to meet such challenges, it is necessary to automate both guidance measures based on text mining and to combine them into a *Grouping and Representing Tool* that is intended to support both students and teachers. To answer the first research question of what text mining methods the tool should be provided with to (1) transform qualitative into quantitative data, (2) to use the quantitative data as a basis to visualize cognitive information and thus increase student's cognitive group awareness, and (3) to further use it as a basis to form groups of heterogeneous learners, we have compared the potentials of various text mining methods in a pre-investigation. We concluded that the extraction of topics and the ascertainment of their extent per student as a basis for visualizing individual cognitive information are done best using LDA combined with Gibbs sampling. The use of LDA for monitoring in a similar educational context was previously suggested by Sherin (2012), and subsequently proved in a similar matter (Southavilay et al. 2013). However, it is important to note that the chosen method is only able to measure learners' quantitative state of knowledge represented in a text and its increase over time. It does not measure qualitative information, such as the number of correct conceptions within a text or revisions of misconceptions, over time.

Furthermore, the Euclidean distance measure could prove its value as an approach to determine differences as a basis for forming heterogeneous groups. The automated group formation according to learners' dissimilarity values corresponds to a comparable grouping as suggested by Manske et al. (2015). They based heterogeneity on different performance characteristics disclosed in students' concept maps. In our case, heterogeneity is solely based on text analysis and primarily defined following the jigsaw schema (cf. Dillenbourg and Jermann 2007): pairs are formed of learning partners with complementary knowledge, which means that learners' expertise per topic is as different as possible from the learning partners' expertise in the same topic. The Euclidian distance

measure cannot assure that the knowledge of two learners is distributed complementarily in each topic, since it operates on the whole text and does not make a distinction between different topics. However, high dissimilarity values guarantee at least a certain degree of complementarity in some topics. For cases in which both learning partners have knowledge to a different extent and provided that this distribution is not one-sided (in the unlikely event that one learner has higher values than the other in all topics), we can also assume that reciprocal role rotation will occur in accordance with the given heterogeneous distribution.

To test the GRT in an experimental study and investigate the second guiding question of the GRT's impact on learning outcomes, we applied the tool in school class with the methods established in the aforementioned pre-investigation. We could find a significant effect of discussion on students' learning. Further, we were able to show that this effect is especially strong when learners use the GRT during their discussion. Beyond that, the use of the GRT was accompanied by a decrease of heterogeneity, or better, a convergence of discussion partners. Since feeding back differences between their states of knowledge encourages discussants to talk about unshared knowledge (Schittekatte and Hiel 1996) and to close knowledge gaps (Sangin et al. 2011), we can conclude that this effect is based on the use of the GRT. This assumption is also supported by the example case of the group with the highest decrease of Euclidian distance after discussion: in half its cases of complementary distribution on a topic (one has written about the topic, the other has not), the one who did not write about the topic within his or her initial text has added content about the respective topic after discussion. Furthermore, we could observe for all cases in which both learning partners had written about a topic to various extents that the learning partner with the visualization of less cognitive information converged to the partners' values. Since the knowledge distribution was not one-sided, this observation supports the assumption that a reciprocal elaboration has taken place. This conclusion is supported by the qualitative analysis of text modifications, since cognitive group awareness encouraged learners to discuss missing topics and to elaborate related concepts only one learner wrote about. Thereby it looks like the topic list provides content that learners add to their text in the case of own knowledge about it. If the visualization shows missing cognitive information, they seem to discuss and exchange related concepts.

It was clear from the beginning that the effect of the GRT on learning could be ascribed to two different functions of the tool, the group formation, and the report of graphical representations. We tolerated this confound to initially clarify that there is an effect utilizing the entire set of features. Finally, the control of variables offered indications weakening this confound, since heterogeneity among learning partners did not differ significantly between the GRT-supported and the unsupported group. This means that the better learning outcomes in the GRT-group could not be attributed to high dissimilarity values, although there was a relationship between Euclidean grouping distance and knowledge acquisition when the whole group was taken into account. Therefore, we assume that the desired structuring of discussion, which according to Dillenbourg and Tchounikine (2007) accompanies heterogeneous group formations, mainly depends on the graphical representation. This conversely means that the effect of the GRT could be even stronger, if we optimize the tool regarding its algorithm for group formation. What remains unanswered is what proportion each component of the graphical representation caused the better learning outcomes, since the representation is characterized by several features: it visualizes a learner's own cognitive information per topic (own bar length), the cognitive information on the learner's teammate per topic (other's bar length), and the cognitive information on the whole group (topic list), which we found very influential as described above. In this context, particularly disentangling effects of providing learners with qualitative or quantitative information needs to be investigated in further research (cf. Erkens et al. 2016).

Despite the necessity for further investigation concerning the aforementioned issues, we can state that, on the whole, the bundling of explicit and implicit guidance mechanisms embodied by the GRT improves collaboration with regard to better structuring of collaborative learning processes and ensures higher learning outcomes for its users than for unsupported learners. Regarding classroom orchestration, the use of the GRT in lessons allows teachers to support their students by guiding their collaboration and to be more efficient at the same time due to the largely automated technology that reduces complexity. Regarding the choice of the underlying computational methods, we have characterized different needs and possible alternative solutions based on approaches that are currently used in similar application fields. Plausible arguments and example-based evidence have been given for these choices. However, this is of course a matter of further investigation.

Further studies should examine the possibilities of putting the learners' focus on content relevant for learning, and fully automating the functions of the GRT which are only semi-automated up to now. Improving the tool especially requires solutions that control for potential sources of errors, e.g. the subjectivity of a teacher or researcher when interpreting clusters, or reduced accuracy and automation while generating the thesaurus and stop words list. It would then be conceivable to apply the findings of this research to learning platforms in order to facilitate online learning processes by recommending learning partners and by increasing students' cognitive awareness.

References

AlSumait, L., Wang, P., Domeniconi, C., & Barbará, D. (2010). Embedding semantics in LDA topic models. In M. W. Berry & J. Kogan (Eds.), *Text mining: applications and theory* (pp. 183–204). Sussex, UK: John Wiley & Sons, Ltd.. doi:10.1002/9780470689646.ch10.

Aronson, E., Blaney, N., Stephan, C., Sikes, J., & Snapp, M. (1978). *The jigsaw classroom*. Beverly Hills, CA: Sage Publications.

Berger, A., Moretti, R., Chastonay, P., Dillenbourg, P., Bchir, A., Baddoura, R., et al. (2001). Teaching community health by exploiting international socio-cultural and economical differences. In P. Dillenbourg, A. Eurelings, & K. Hakkarainen (Eds.), *Proceedings of the first European Conference on Computer Supported Collaborative Learning* (pp. 97–105) .Retrieved from http://www.eculturenet.org/mmi/euro-cscl/Papers/14.pdf

Berlyne, D. (1966). Notes on intrinsic motivation and intrinsic reward in relation to instruction. In J. Bruner (Ed.), *Learning about learning (Cooperative Research Monograph No. 15)*. Washington, DC: Department of Health, Education, and Welfare, Office of Education.

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM, 55*(4), 77–84. doi:10.1145/2133806.2133826.

Blei, D. M., Ng, A. Y., & Jordan, I. J. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research, 3*, 993–1022. doi:10.1162/jmlr.2003.3.4-5.993.

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.

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., & Scholvien, A. (2014). *Providing knowledge-related partner information in collaborative multimedia learning: Isolating the core of cognitive group awareness tools*, In Proceedings of the 22nd International Conference on Computers in Education ICCE 2014 (pp. 171–179). Japan: Nara.

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.

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.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science, 41*(6), 391–407.

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.

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, NL: Open Universiteit Nederland.

Dillenbourg, P., & Betrancourt, M. (2006). Collaboration load. In J. Elen & R. E. Clark (Eds.), *Handling complexity in learning environments: research and theory* (pp. 142–163). Amsterdam, NL: 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). Springer US.

Dillenbourg, P., & Tchounikine, P. (2007). Flexibility in macro-scripts for computer-supported collaborative learning. *Journal of Computer Assisted Learning, 23*(1), 1–13. doi:10.1111/j.1365-2729.2007.00191.x.

Dillenbourg, P., Järvelä, S., & Fischer, F. (2009). The evolution of research on computer-supported collaborative learning. In D. N. Balacheff, D. S. Ludvigsen, D. T. de Jong, D. A. Lazonder, & D. S. Barnes (Eds.), *Technology-Enhanced Learning (pp. 3–19)* .Springer Netherlands

Dimitriadis, Y. A. (2012). Supporting teachers in orchestrating CSCL classrooms. In A. Jimoyiannis (Ed.), *Research on e-Learning and ICT in Education (pp. 71–82)*. Springer .Retrieved from http://link.springer.com/chapter/10.1007/978-1-4614-1083-6_6

Doise, W., & Mugny, G. (1984). *The social development of the intellect*. (W. Doise & G. Mugny, Eds.) (Vol. 10). Amsterdam, Netherlands: Pergamon. Retrieved from http://www.sciencedirect.com/science/article/pii/B978008030215750010X.

Engelmann, T., & Hesse, F. W. (2010). How digital concept maps about the collaborators' knowledge and information influence computer-supported collaborative problem solving. *International Journal of Computer-Supported Collaborative Learning, 5*(3), 299–319. doi:10.1007/s11412-010-9089-1.

Erkens, M., Schlottbom, P., & Bodemer, D. (2016). Qualitative and quantitative information in cognitive group awareness tools: Impact on collaborative learning. In U. Cress (Ed.), *Transforming learning, empowering learners: the international conference of the learning sciences (ICLS) 2016*. International Society of the Learning Sciences: Singapore.

Feinerer, I., & Hornik, K. (2013). Tm: text mining package. A framework for text mining applications within R. Retrieved from http://tm.r-forge.r-project.org/.

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.

Flavell, J. H., & Botkin, P. T. (1968). The development of role-taking and communication skills in children. Wiley.

Graesser, A. C., Lu, S., Jackson, G. T., Mitchell, H. H., Ventura, M., Olney, A., & Louwerse, M. M. (2004). AutoTutor: A tutor with dialogue in natural language. *Behavior Research Methods, Instruments, & Computers, 36*(2), 180–192. doi:10.3758/BF03195563.

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. Haake (Eds.), *Scripting computer-supported communication of knowledge-cognitive, computational and educational perspectives* (pp. 91–98). New York, NY: Springer.

Hung, J. (2012). Trends of e-learning research from 2000 to 2008: Use of text mining and bibliometrics: Research trends of e-learning. *British Journal of Educational Technology, 43*(1), 5–16. doi:10.1111/j.1467-8535.2010.01144.x.

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. (1970). *The social psychology of education*. New York, NY: Holt, Rinehart and Winston.

Johnson, D. W. (1980). Group processes: influences of student-student interaction on school outcomes. In J. McMillan (Ed.), *The social psychology of school learning* (pp. 123–168). New York, NY: Academic Press.

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.

Johnson, D. W., Johnson, R. T., & Smith, K. A. (1998). Cooperative learning returns to college what evidence is there that it works? *Change: The Magazine of Higher Learning, 30*(4), 26–35. doi:10.1080 /00091389809602629.

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*(2–3), 211–224. doi:10.1007/s11412-007-9014-4.

Kohlberg, L. (1969). *Stage and sequence: the cognitive-developmental approach to socialization.* Chicago, IL: Rand McNally.

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.

Kreijns, K., Kirschner, P. A., & Jochems, W. (2003). Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: a review of the research. *Computers in Human Behavior, 19*(3), 335–353. doi:10.1016/S0747-5632(02)00057-2.

Lee, Y., & Nelson, D. (2004). Instructional use of visual representations of knowledge. In R. Ferdig, C. Crawford, R. Carlsen, N. Davis, J. Price, R. Weber, & D. Willis (Eds.), *Proceedings of Society for Information Technology & Teacher Education International Conference 2004* (pp. 2371–2378). Chesapeake, VA: Association for the Advancement of Computing in Education (AACE).

Leydesdorff, L., & Hellsten, I. (2006). Measuring the meaning of words in contexts: An automated analysis of controversies about "Monarch butterflies," "Frankenfoods," and "stem cells.". *Scientometrics, 62*(2), 231–258. doi:10.1556/Scient.67.2006.2.6.

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

Manske, S., Hecking, T., Chounta, A., Werneburg, S., & Hoppe, H. U. (2015). Using differences to make a difference: A study on heterogeneity of learning groups. In O. Lindwall, P. Häkkinen, T. Koschmann, P. Tchounikine, & S. Ludvigsen (Eds.), *Exploring the material conditions of learning: the computer supported collaborative learning (CSCL) conference 2015* (Vol. 1, pp. 182–189). Gothenburg: The International Society of the Learning Sciences.

Miller, T. W. (2005). *Data and text mining: a business applications approach.* Upper Saddle River, N.J: Prentice Hall.

Miner, G., Elder, J., Fast, A., Hell, T., Nisbet, B., & Delen, D. (2012). *Practical text mining and statistical analysis for non-structured text data applications.* Boston: Academic Press.

O'Donnell, A. M., & O'Kelly, J. (1994). Learning from peers: Beyond the rhetoric of positive results. *Educational Psychology Review, 6*(4), 321–349. doi:10.1007/BF02213419.

Piaget, J. (1950). *The psychology of intelligence.* New York, NY: Harper.

Porter, M. F. (1980). An algorithm for suffix stripping. *Program, 14*(3), 130–137.

Rajman, M., & Vesely, M. (2004). From text to knowledge: Document processing and visualization: A text mining approach. In S. Sirmakessis (Ed.), *Text mining and its applications* (Vol. 138, pp. 7–24). Springer: Berlin, Heidelberg. doi:10.1007/978-3-540-45219-5_2.

Rosé, C., Wang, Y.-C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., & Fischer, F. (2008). Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. *International Journal of Computer-Supported Collaborative Learning, 3*(3), 237–271. doi:10.1007/s11412-007-9034-0.

Ryan, R. M., & Deci, E. L. (2006). Self-regulation and the problem of human autonomy: Does psychology need choice, self-determination, and will? *Journal of Personality, 74*(6), 1557–1586. doi:10.1111/j.1467-6494.2006.00420.x.

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. (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.

Sherin, B. (2012). Using computational methods to discover student science conceptions in interview data. In *LAK '12 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 *Proceedings of the third international conference on learning analytics and knowledge* (pp. 38–47). New York, NY: ACM. doi:10.1145/2460296.2460307.

Springer, L., Stanne, M. E., & Donovan, S. S. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. *Review of Educational Research, 69*(1), 21–51. doi:10.3102/00346543069001021.

Suthers, D. (2001). Towards a systematic study of representational guidance for collaborative learning discourse. *Journal of Universal Computer Science, 7*(3), 254–277.

Suthers, D., & Hundhausen, C. D. (2003). An experimental study of the effects of representational guidance on collaborative learning processes. *The Journal of the Learning Sciences, 12*(2), 183–218. doi:10.1111/3.5739.

Tane, J., Schmitz, C., & Stumme, G. (2004). Semantic resource management for the web: An e-learning application. In *Proceedings of the WWW conference* (pp. 1–10). New York, NY: ACM.

Villalon, J., Kearney, P., Calvo, R. A., & Reimann, P. (2008). Glosser: Enhanced feedback for student writing tasks. In *In Eighth IEEE International Conference on Advanced Learning Technologies, 2008. ICALT '08* (pp. 454–458). Santander, Cantabria: IEEE. doi:10.1109/ICALT.2008.78.

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. doi:10.1007/978-0-387-36949-5_12 .