Computers & Education 90 (2015) 80–94

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Teacher regulation of cognitive activities during student collaboration: Effects of learning analytics

Anouschka van Leeuwen*, Jeroen Janssen, Gijsbert Erkens, Mieke Brekelmans

Utrecht University, Department of Education, Heidelberglaan 1, 3584 CS Utrecht, The Netherlands

ARTICLE INFO

Article history:
Received 9 April 2015
Received in revised form 10 September 2015
Accepted 11 September 2015
Available online 14 September 2015

Keywords:
Cooperative/collaborative learning
Secondary education
Teaching/learning strategies

ABSTRACT

By collaboratively solving a task, students are challenged to share ideas, express their thoughts, and engage in discussion. Collaborating groups of students may encounter problems concerning cognitive activities (such as a misunderstanding of the task material). If these problems are not addressed and resolved in time, the collaborative process is hindered. The teacher plays an important role in monitoring and solving the occurrence of problems. To provide adaptive support, teachers continuously have to be aware of students' activities in order to identify relevant events, including those that require intervention. Because the amount of available information is high, teachers may be supported by learning analytics. The present experimental study ($n = 40$) explored the effect of two learning analytics tools (the Concept Trail and Progress Statistics) that give information about students' cognitive activities. The results showed that when teachers had access to learning analytics, they were not better at detecting problematic groups, but they did offer more support in general, and more specifically targeted groups that experienced problems. This could indicate that learning analytics increase teachers' confidence to act, which in turn means students could benefit more from the teacher's presence.

1. Introduction

Computer-supported collaborative learning (CSCL) is an instructional strategy in which collaboration among students is supported by technology. It is based on the idea that collaboration is beneficial for learning. By collaboratively solving a task, students are challenged to share ideas, express their thoughts, and engage in discussion (Stahl, Koschmann, & Suthers, 2006). Learning during CSCL is seen as an interactive, constructive, and largely self-regulated process. Students' learning activities can be categorized into cognitive activities (i.e., related to the content of the task, for example structuring and analyzing task material), social activities (for example, the occurrence of discussion in terms of agreement and disagreement and participation rates of group members), and regulative activities at both the cognitive and social level (for example, discussing strategies for solving the task) (Janssen, Erkens, Kanselaar, & Jaspers, 2007; Kaendler, Wiedmann, Rummel, & Spada, 2014; Vermunt & Verloop, 1999; Weinberger & Fischer, 2006).

Digital learning environments designed for collaborative learning generally integrate tools for carrying out the task as well as for communication between group members. Together, these tools facilitate the types of student activities mentioned above because they support the sharing of resources and provide an opportunity for communication within the group

* Corresponding author.
E-mail address: A.vanLeeuwen@uu.nl (A. van Leeuwen).

http://dx.doi.org/10.1016/j.compedu.2015.09.006

(Erkens, Jaspers, Prangsma, & Kanselaar, 2005). In the present study, collaboration occurs through a digital learning environment in which students have access to task materials, share a text-editor with their group members, and communicate through a chat facility. Providing these tools, however, does not guarantee that students will adequately finish their task, nor a high quality of discussions (Kirschner & Erkens, 2013; Pargman, 2003; Rummel & Spada, 2005). During CSCL, teachers act as a facilitator of students' activities (Kaendler et al., 2014). Teachers can for example offer thoughts that deepen or broaden the discussion and keep track of the progress that groups of students are making on the task. To do so, it is important that teachers are able to identify all relevant events, including those that require intervention. Because of the generally rapid pace of activities within synchronous CSCL settings and the large amount of available information, supporting student activities is a demanding task. In the present study, we focus on teacher regulation of groups' cognitive activities, which are important because they are directly related to for example knowledge acquisition and of which it is known that students may experience problems (Weinberger & Fischer, 2006). We explore a way of supporting the teacher, namely by visualizations of the collaborating groups' activities. The sections below describe students' cognitive activities, the teacher's role during CSCL, and how the teacher may be supported while regulating students' cognitive activities.

1.1. Students' cognitive activities during collaborative writing tasks

The present article is situated in the context of a collaborative writing task. Groups of students in secondary education synchronously communicate with each other through a chat tool and share a text editor to write an essay based on historical sources, which are all provided within the learning environment. The cognitive activities involved in this task include evaluating and discussing the task material, writing the essay, and reading historical sources. At the level of regulative activities, the groups have to agree on a strategy for completing the task and to monitor their progress. As stated before, students largely self-regulate their activities, but it is known that problems may occur that could negatively influence students' learning gains or the quality of the group product. Two of those problems are described in this section.

The first problem concerns discussion of task material within groups. Researchers generally distinguish between on-task and off-task communication within group discussions (see De Wever, Schellens, Valcke, & Van Keer, 2006, for a review). Discussing the content of the task is most clearly related to knowledge acquisition (Weinberger & Fischer, 2006; Cohen, 1994; see also Carroll's Time-On-Task hypothesis, Carroll, 1963, quoted in Baker, Corbett, Koedinger, & Wagner, 2004). Because of the informal character of synchronous chat communication, students may stray off-task during discussions, which could lead to decreased learning gains. When groups do stay on-task, there is another potential difficulty, namely that the discussion has insufficient breadth (Baker, Andriessen, Lund, Van Amelsvoort, & Quignard, 2007). That is, discussions may be superficial or one-sided when the topic of the discussion lingers on only a limited set of the concepts that are relevant to the task. Limited breadth of discussion could also mean less depth, because the students did not take into account all possible explanations or viewpoints and did not connect these views to each other (Baker et al., 2007). So, the content of group discussions, in terms of on- and off-task behavior and the concept coverage or breadth of the discussion, is one cognitive aspect that teachers can help students to regulate.

The second problem is concerned with how students alternate between cognitive activities. While solving the task, students continuously alternate between writing and discussing (Rummel & Spada, 2005), and engage in activities such as outlining, composing, and reviewing the written text. The groups of students may choose different strategies for writing, such as parallel exploration of the material followed by integration of ideas, or continuous joint construction of text (Onrubia & Engel, 2009). For all strategies, it is important that time is managed in an adequate way. Groups may thus need help to monitor their progress while they engage in the multiple cognitive activities involved with collaborative writing.

If these problems are not addressed and resolved in time, the collaborative process is hindered. The teacher plays an important role in monitoring and solving the occurrence of problems as will be explained below.

1.2. Teacher regulation of students' cognitive activities

The change toward the use of collaborative learning in education also requires changes on the part of teachers. In case of CSCL, teacher regulation takes shape by monitoring the learning activities of students as they independently work with other students on their group assignments, and intervening with feedback and assistance when needed (Anderson, Rourke, Garrison, & Archer, 2001; Kaendler et al., 2014). When the educational goals are to analyze, evaluate, and synthesize knowledge, leading the students towards interaction and experimentation, teacher regulation is more loose (Salinas, 2008; Vermunt & Vermetten, 2004). Even though there is more loose teacher regulation during CSCL, the teacher maintains an important role (Kaendler et al., 2014). One of the teacher's tasks is to monitor the occurrence of problems and to help to resolve them. When problems arise or students do not make enough progress, teachers can offer their assistance. Many researchers have tried to analyze the effects of teaching activities on learning outcomes or the quality of group products during CSCL (for example Hsieh & Tsai, 2012 and Onrubia & Engel, 2012, see Van Leeuwen, Janssen, Erkens, & Brekelmans, 2013, for an overview). From these studies, it appears that the effectiveness of teaching is largely determined by the adaptivity (content and timing) of teacher interventions (Gibbs & Simpson, 2004). Each group of students has different needs, to which the teacher should adapt (Van Leeuwen et al., 2013; Coll, Rochera, & de Gispert, 2014). Thus, as a result of correctly timed and correctly chosen interventions, teacher regulation of CSCL can effectively help collaborating groups.

To provide adaptive support, teachers have to continuously be aware of students' activities in order to identify relevant events, including those that require intervention (Kaendler et al., 2014). The teacher has to maintain an overview of events, a characteristic that in subsequent sections will be called a diagnosis of the current situation. From the previous description of the cognitive activities, it becomes apparent that diagnosing those activities is a demanding task. There are multiple groups to monitor, and to work on the group assignment each of them engages in multiple activities in multiple tools. The pace of the groups' discussions can be fast. Like regular classrooms, the simultaneity of the situation (Doyle, 2006) means that there is a large amount of information to tend to. This may lead teachers to experience high cognitive load, i.e., diagnosing multiple groups of students demands a large amount of cognitive capacity (Paas, Tuovinen, Tabbers, & Van Gerven, 2003). When cognitive load is too high, this may cause teachers to fall back on automated courses of action, which means accurate and up-to-date diagnosing of student performance does not occur (Feldon, 2007). A possible consequence is that teachers use their existing knowledge about students to decide on an intervention instead of using information about the current situation (Feldon, 2007; Schwarz & Asterhan, 2011). This could mean that the intervention is not optimally aligned with students' needs. It is therefore expected that an increase in cognitive load could lead to less adaptation to students' needs.

1.3. Supporting teachers with learning analytics

To lower cognitive load, teachers may be supported by the addition of learning analytics tools. At the intersection of technical analysis of data and the learning sciences, learning analytics (LA) is defined as "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs" (Siemens & Gasevic, 2012). In the present case, this means that teaching may be supported by analyzing student activities and reporting these analyses back to the teacher. When student activities are summarized and visualized in an easily understandable way, this could lower the time and effort needed by teachers to monitor all groups of students' activities. The LA can enhance the teacher's knowledge or awareness of each groups' activities (De Groot et al., 2007; Voyiatzaki & Avouris, 2014).

A large body of research has focused on supporting *students* with tools that enhance students' awareness of the activities of their group members. So called group awareness tools can provide students with information about social or cognitive characteristics of the group such as knowledge, attention and attitudes of each group member (Buder & Bodemer, 2008). These awareness tools may guide students in their collaboration by stimulating discussion of the displayed information (Dehler, Bodemer, Buder, & Hesse, 2011) and by giving the groups of students feedback about the state of the collaboration concerning a range of indicators (Harrer, McLaren, Walker, Bollen, & Sewall, 2006; Jermann, Soller, & Muehlenbrock, 2005).

Partly, the functions that group awareness tools fulfill for students are also applicable to the teacher, because both students and teachers can use them to monitor the students' activities (Verbert et al., 2014). However, such tools can also have a function specifically for either students or teachers. For example, consider a LA tool that visualizes and compares a characteristic of collaboration in each group (for example, the number of solved tasks). Students and teachers will derive information from this tool for different purposes. Shown to the students, such an overview for example creates opportunities for social comparison, which might motivate students to set higher standards for themselves (Janssen et al., 2007; Michinov & Primois, 2005). The teacher, on the other hand, uses the information to maintain an overview of the class and possibly as an indication of which groups might need additional help. As stated, this way the LA can contribute to lowering the cognitive load associated with monitoring the activities of multiple groups at the same time. Many articles have described how such tools that support teachers may be designed, making use of techniques such as data mining to discover patterns between student activities in learning environments and their resulting learning gains (cf. Jermann et al., 2005; Romero & Ventura, 2010). However, the step of actually providing teachers with these tools and, consequently, studying how they are used, is rarely taken (Chatti, Dyckhoff, Schroeder, & Thüs, 2012; Papamitsiou & Economides, 2014). A few exceptions are described here. Dyckhoff, Zielke, Bültmann, Chatti, and Schroeder (2012) developed diagnostic tools in collaboration with teachers. Teacher interviews revealed that the tools were perceived as being useful, especially because it allowed teachers to observe whether changes in learning materials or teaching strategies led to changes in student behavior. Mazza and Dimitrova (2007) also created diagnostic tools, and besides interviewing teachers also performed a small scale experimental study (with 6 participants) which showed that teachers could more quickly and more accurately gain an overview of students' activities when they used the tools. Casamayor, Amandi, and Campo (2009) found that when a teacher was notified of possibly problematic situations during students' collaboration, the teacher intervened more and was able to solve more conflicts. A qualitative study that focused both on diagnosing and intervening was conducted by Schwarz and Asterhan (2011). These authors described the choices of a single teacher concerning diagnosis and interventions and how the LA tools support the teacher to achieve adaptivity ("approximate attunement", p. 436) to the needs of each group. In particular, the tools gave the teacher an overview of activities, which resulted in the mental space needed to carefully consider the appropriate intervention at the appropriate time. Similarly, Chounta and Avouris (2014) found that as class size increased, teachers had more need of LA to retain overview. After teachers focused on a particular group or on an individual student, the LA helped them to regain an overview of the rest of the class. Thus, LA seem to be a promising direction for supporting teachers during regulation of synchronous CSCL. However, research in this area is still scarce and primarily small scale. There is thus a need for further systematic examination of the effects of LA on teacher regulation of CSCL. The goal of this article is to add to this knowledge base, specifically focusing on the effects of LA on teacher diagnosing and intervening. The effects of two learning analytics tools are studied that provide teachers with information about students' cognitive activities. Two teacher supporting tools were added to the existing collaborative learning environment called VCRI (explained in more detail in the method section).

1.4. Learning analytics used in this study

The first tool, called the Concept Trail (CT), is a timeline that displays the topics that groups are discussing in the Chat window. The CT marks the occurrence of a predefined set of task-related concepts and their synonyms by putting dots on the timeline. The development of the topics in the group discussion therefore becomes visible at a glance. As the group discussion progresses, older Chat messages will move upwards out of sight, but the CT still displays the entire history. The CT is integrated into the existing Chat windows, so that the development of the discussion becomes visible each time the teacher looks at a group's discussion.

Fig. 1 shows a screenshot of a Chat window with the integrated CT on top. On the left side of the CT, five concepts are displayed. The dots on the timeline indicate that the concept is mentioned in the Chat conversation. Furthermore, in the conversation itself the concepts are highlighted in yellow, so that it is immediately visible in what part of the message the concept was used. The grey bar within the CT corresponds to the current timeframe of the Chat conversation and can be moved by the teacher so that the teacher can easily jump to another episode within the conversation.

The CT gives information about cognitive activities in two ways. First of all, it can inform the teacher whether a group is on-task or not. As was described earlier, the more students are on-task, the more likely they are to acquire knowledge and to engage in meaningful discussion. Secondly, the teacher can not only monitor how often concepts are mentioned, but also the concept coverage of the discussion. For example, it might be the case that the group only covers one or two important aspects of the task, and that there are no dots on the timeline for the other concepts. It could also be the case that dots are displayed for multiple concepts within the same time frame, which could indicate that students try to connect concepts to each other. The CT thus allows the teacher to monitor the discussion and to give suggestions to deepen or broaden the discussion by mentioning specific concepts. The resulting increased specificity of interventions is desirable because it helps learners to understand how well they are performing and what still needs to be accomplished (Voerman, Meijer, Korthagen, & Simons, 2012).

The second teacher supporting tool is the availability of Progress Statistics (PS), which informs teachers of the progress of each group in terms of the number of written words in the text editor and the chat tool. Fig. 2 displays a screenshot of the PS of the collaboratively written texts. The five bars represent the progress of the five groups concerning the number of words they have written in the text editor. The horizontal line indicates the class average. The same type of PS are available for the number of words posted in the chat conversation, which can be selected from the menu on the left.

The PS give teachers an indication of the amount of activity within the groups. The class average enables teachers to see at a glance whether there are any groups that show deviating amounts of activity, i.e., groups that are behind as well as those that have more written text or more chat utterances than the others. Also, teachers may compare the activity in the chat with the activity in the text editor for each group, thereby monitoring whether groups balance their activities between writing and discussing.

It is important to note two things about the teacher supporting tools. First of all, the tools on their own do not give a complete picture of all groups' cognitive activities. The quality of the written texts, for example, does not necessarily correlate with the amount of words written. Neither does mentioning a concept mean that a student understands its content. Together, however, the tools give an aggregated overview of the situation. Although the tools are based on relatively simple measures, the information shown can be an indication for the teacher whether more time needs to be spent on a particular group. This brings us to the second point. The supporting tools do what their name suggests: support the teacher, but not replace or decide for the teacher. Whether intervention is necessary remains to the teacher to decide.

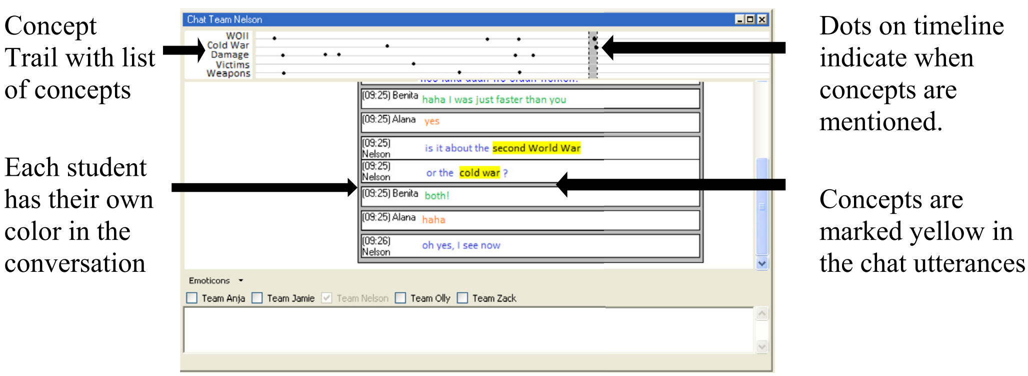

Fig. 1. Screenshot of the Chat Tool with integrated Concept Trail (translated from Dutch).

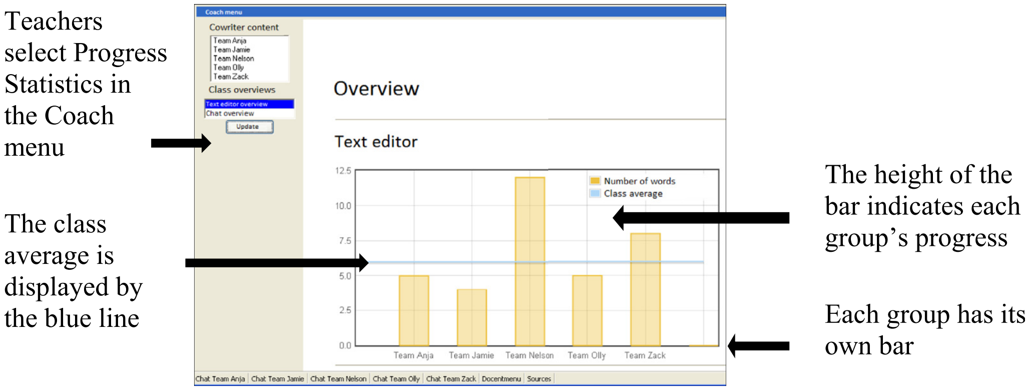

Fig. 2. Screenshot of the Progress Statistics tool (translated from Dutch).

1.5. Research questions

The following research questions were formulated:

What is the effect of teacher supporting tools that show information on concept coverage in discussions and task progress of collaborating groups on:

1) Teachers' diagnosis of concept coverage within discussions and the task progress of collaborating groups?
2) The frequency, focus, and specificity of teacher interventions?
3) Teachers' self-reported cognitive load?

2. Method

2.1. Design

An experimental study with a between and within subjects design was conducted to test the effects of the Concept Trail and the Progress Statistics. Participants were randomly assigned to the control (no LA tools) or the experimental (LA tools) condition. In both conditions, participants went through vignettes of three collaborative situations. The independent variable is therefore the absence or presence of LA tools. The dependent variables are the participants' diagnosis, interventions and cognitive load during and after each of the vignettes, which constitute three repeated measures.

2.2. Participants

Forty undergraduate educational sciences students participated in this study as a part of the university course they were enrolled in concerning the subject of ICT use in education. The educational sciences students were asked to act as teachers during collaborative situations (see below). The participants were considered to be able to represent teachers because they have background knowledge about education and affinity and experience with educational practice.

Participants were randomly divided over the control and experimental condition, resulting in 20 participants in each condition (3 males and 17 females each), with a mean age of 21.1 years ($SD = 3.9$) and 19.9 years ($SD = 1.4$) respectively.

2.3. Simulation software and teacher supporting tools

To combine an experimental setup with the use of authentic situations, simulation software was created based on an existing CSCL environment called VCRI (Virtual Collaborative Research Institute, see for example Van Leeuwen et al., 2013). The simulation software is able to read student data from previous research and to display it as if students are collaborating real time. This way, all participants in the current study could be shown the same authentic collaborative situations. At the same time, an experimental set-up was realized by creating two versions of the simulation software: the control version, which was essentially the same as VCRI, and the experimental version, which was enhanced with teacher supporting tools. The same set-up was used in an earlier study (Van Leeuwen, Janssen, Erkens, & Brekelmans, 2014).

In VCRI, students work on collaborative writing assignments. They communicate through a Chat tool and work on the assignment in a shared text editor, the Cowriter. In the control condition of the simulation software, participants were able to follow all students' activities. They could open Chat windows and read the task materials that were available to students (Sources). Through the Coach menu, participants could check the texts that are being written in the Cowriter. Besides these ways to diagnose, participants could intervene by sending messages in the Chat windows. Multiple windows can be opened simultaneously, allowing the teacher to monitor multiple groups and their activities in multiple tools at the same time. The control condition was thus similar to the way VCRI is normally used in educational settings. The synchronous, digital communication offers several advantages. For example, it ensured that the teacher could constantly monitor each group's activities and lowered the chance of classroom disturbances or interruptions (Petrou & Dimitracopoulou, 2003). Digital communication through a Chat-tool can also be argued to lead to more precise articulation of ideas because of the absence of non-verbal cues (Asterhan & Schwarz, 2010) and engages both active and more silent students to participate (Asterhan & Eisenmann, 2011).

In the experimental condition, the software from the control condition was enhanced to include the CT and PS (described in the theoretical introduction). Thus, whereas the control condition also had access to students' Chat Tools and their written essays, the participants in the experimental condition had additional access to the CT and the PS.

2.4. Vignettes

Each participant went through vignettes of three collaborative situations. In our earlier, similar study in which *four* vignettes were presented to the participants (Van Leeuwen et al., 2014), relatively small effects of vignette order were found. Therefore, the three vignettes in the present study were presented in a fixed order. Each vignette contained five collaborating groups and lasted for 8 min, during which student activities occurred. The situations were selected from existing data from previous research in secondary education, in which VCRI was used by students and their teachers to work on collaborative assignments. The students worked on an assignment concerning the Second World War and the Cold War at the level of secondary education, of which it was expected that all participants had at least basic knowledge. The assignment asked students to examine the consequences of the Second World War and the causes for and development of the Cold War. They were asked to reflect on reasons why a Third World War could have happened, and why in the end it did not. Multiple historical sources were provided, including maps, speeches by historical figures, and explanations of important concepts. The students had to collaboratively evaluate and discuss these sources, integrate their meaning into the historical context, and finally, to write an argumentative essay about the main questions listed above. All communication between students occurred digitally by communicating through a chat tool.

By replaying students' activities in the simulation software, participants were able to act as the teacher as if the situation was happening real time. Chat messages appeared, the texts in the Cowriter changed, and in case of the experimental condition the supporting tools were constantly updated to reflect the current situation. Participants could open and close the tools of multiple groups at the same time and thus decide which groups they would monitor at any given moment. All names were changed to preserve anonymity of students. Small changes were made in the students' Chat messages, for example when students referred to the teacher.

Three vignettes were selected to reflect the cognitive aspects that the teacher supporting tools are aimed at: concept coverage within the discussions and task progress. The first vignette contained a group that had low task progress. The second vignette contained one group that had very high task progress because they copied text from the sources, and one group that had low concept coverage in their discussion. The third vignette contained a group with low concept coverage in their discussion. Table 1 summarizes the problems that occurred in the vignettes. Thus, each vignette contained one or two problematic groups and three or four non-problematic groups (i.e., groups that did not experience problems concerning cognitive activities).

2.5. Procedure

During the experiment, participants could ask the researcher questions if they were unsure about how to use the software. In total, the task took participants 1 h. Participants went through the following steps:

1. Participants were handed a 1-page summary of the Cold War to serve as background information to the task that the students in the vignettes worked on. After the participant had read this page, they preceded to step 2. The summary was available throughout the experiment.

Table 1
Problem descriptions of the four vignettes, in order of appearance.

Vignette	Characterization	Problem description
(1)	Low task progress	One group of students is far behind the others concerning progress on the task. The concept coverage in the discussion is high, but they do not transfer their thoughts to the text editor.
(2)	High task progress	One group of students has a lot more written text than the others, because they have simply copied and pasted information from the task materials.
	Low concept coverage	The concept coverage of the discussion of one of the groups is low, caused by disagreements among the group members.
(3)	Low concept coverage	The concept coverage of the discussion of one of the groups is low; this group has a lot of off-task activity.

2. An introductory video was shown in which an explanation was given of the procedure of the experiment and of how to use the software. The participant then logged on to the simulation software. The program went through several steps.
3. A test vignette was played in which participants could practice opening windows and typing messages, and thus get used to the software. This test vignette lasted 2 min.
4. The three experimental vignettes were offered one by one (see Table 1). Each vignette was preceded by a brief explanation of the situation the vignette was taken from, i.e., what subject the class was working on and how many lessons they had worked on the topic. Then the vignette played, lasting 8 min, during which the participants were free to type messages to the students and to check their progress. Participants were explicitly asked in the introductory movie to act as they would in a real educational setting, even though in this case the students could not respond to the participants' messages. At the end of each vignette, a screen opened with questions about the participant's experienced cognitive load, and their diagnosis of task progress and discussion within the groups.
5. A final short questionnaire was administered in which participants were asked about the set-up of the experiment. On a scale from 1 (disagree) to 5 (agree), participants on average scored 4.41 ($SD = 0.79$) for the question whether the set-up of the experiment was clear after viewing the introduction video and playing the test vignette.

2.6. Instruments

The participants' actions were automatically logged during and after each vignette. This resulted in the following types of data. First of all, it was logged what tools the participant used and how often, for example opening a group's Chat window or selecting information in the Coach menu (diagnosis). Also, all messages sent by the participants to one of the groups were automatically collected (interventions). After each vignette, participants gave scores to each of the groups, and had the option to give an explanation of the given scores (overall diagnosis). Furthermore, after each vignette participants were asked to give a score to their experienced cognitive load.

2.6.1. Diagnosis during vignettes

During the vignettes, the following actions were logged: for the Chat tool, whether participants opened a group's Chat screen and whether they scrolled through the Chat history; for the Coach menu, whether participants opened this menu and which information they selected for which group; and for the Sources tool, whether participants opened this tool and opened sources in order to read them.

2.6.2. Interventions

During each vignette, participants were able to intervene by sending messages in a group's Chat-tool. We were interested in the effect of the learning analytics on the participants' attention to the cognitive aspects shown by the tools and what kind of messages they would send to the groups. In other words, we wanted to know what learning activity the interventions were aimed at (focus) and what type of regulation (means) participants chose (Van Leeuwen et al., 2013).

To analyze the *focus* of interventions we included four categories for the two main cognitive activities students perform to work on the group assignment, namely discussing and writing. To be able to analyze the prevalence of a focus on cognitive aspects, we added categories for interventions concerning students' social activities and general task strategy (see Table 2).

To analyze the *means* of interventions we included categories to reflect the degree of teacher regulation as well as how participants evaluated the activities of students. For example, an "instruction" signifies stronger regulation than a

Table 2
Coding scheme for interventions, consisting of focus, means, and specificity.

Dimension	Category	Explanation
Focus/learning activity	Discussion on/off task	Comments about students being on-task or off-task
	Discussion content	Comments about the content of the discussion, for example its correctness and concept coverage
	Text quality	Comments about the quality of the written texts, for example use of arguments or use of sources
	Text appearance	Comments about the appearance of the written text, for example the lay-out, amount of text, and use of headers
	Strategy	Comments about general task strategy
	Social activities	Comments about students' social activities, for example group participation and occurrence of conflict
	Other	All remaining interventions
Means/type of regulation	Diagnosing	Asking about a group's progress or current situation
	Positive evaluation	A positive judgment of a student's behavior or of group products
	Suggestion	Giving suggestions for the discussion or for writing the text
	Explanation	Answering questions or giving explanations
	Instruction	Giving a direct order or instruction
	Negative evaluation	A negative judgment of a student's behavior or of group products
	Other	All remaining interventions
Specificity	Concepts	Mentioning concrete concepts from the discussion or the written text. Also includes quoting chat messages
	Other	All remaining interventions

"suggestion", and "negative or positive evaluations" are more outspoken judgments than neutral interventions such as "explanations" or "suggestions".

Besides the focus and means, interventions were also coded for *specificity*. It was hypothesized that the availability of the CT would help participants to give more specific feedback. Therefore, interventions were coded whether or not they contained specific mention of task-related concepts.

Table 2 displays the full list of categories that were used for focus, means and specificity. Two researchers independently coded a sample of 100 interventions, resulting in Cohen's Kappa of 0.77 (focus), 0.81 (means) and 0.95 (specificity). These interrator reliabilities were interpreted as good and the remaining interventions were coded by one of the researchers.

2.6.3. Overall diagnosis

Overall diagnosis was measured in two ways. First of all, after each vignette participants were asked to give a score to each group for the concept coverage in the group's discussion and for the group's task progress on a scale from 1 to 10 (where 1 = very poor and 10 = excellent). Participants could also indicate they had not been able to make a proper judgment and refrain from giving a score. In total, each vignette resulted in ten scores, two for each of the five groups.

As a second measure of overall diagnosis participants could write a comment to explain their given score. These comments were collected and coded. Three aspects were of interest to answer our research question. First of all, the focus of the comments was coded, which indicates what aspect of students' activities was mentioned to explain a given score. To code the focus of the comments, the coding scheme for the focus of interventions was used, see Table 2. Besides the focus, it was coded what source of information participants used to come up with their score, i.e., which tool participants reported as the source of their judgment. Lastly, the specificity of the judgment was coded, as a measure of how elaborate the explanation of the diagnosis was. Besides mentioning specific concepts (see Table 2), specificity could also mean that groups were compared or that participants made a reference to the development within a group. Table 3 displays the categories that were used. A sample of 100 comments was independently coded by two researchers, resulting in Cohen's Kappa of 0.84 (focus), 0.87 (source) and 0.79 (specificity). These interrater reliabilities were interpreted as good and the remaining comments were coded by one researcher.

2.6.4. Cognitive load

To measure cognitive load, participants were asked after each vignette to indicate the amount of effort it took them to complete the task. Following Paas et al. (2003), mental effort is defined as "the aspect of cognitive load that refers to the cognitive capacity that is actually allocated to accommodate the demands imposed by the task; thus, it can be considered to reflect the actual cognitive load" (p. 64). Using the 9-point symmetrical category mental effort rating scale (Paas et al.,2003), participants were asked to indicate the amount of effort it took them to complete the task on a scale from 1 to 9, where 1 indicates very low effort and 9 means very high effort. The following question was asked: 'How much effort did it take to maintain an overview of the collaborating groups?'. For each participant, three measurements of cognitive load were obtained.

3. Results

First, information is given about the usage of the supporting tools (the Concept Trail (CT) and the Progress Statistics (PS)). Then, each research question is answered.

3.1. Use of supporting tools

Table 4 shows the average number of actions per participant per tool per vignette. The CT was integrated in each of the groups' Chat windows. Participants in the experimental condition did not open or focus on the Chat windows more often than participants in the control condition ($t(38) = -0.64$, $p = .53$, $d = 0.21$). Participants could use the scroll bar to view the Chat history. Although the average amount of scrolling is twice as high in the control condition, this difference is not significant ($t(38) = 1.32$, $p = .15$, $d = 0.47$).

Table 3
Coding scheme for the source and specificity of comments of overall diagnosis scores.

Dimension	Category	Explanation
Source	Progress statistics	Explicitly mentioning the Progress Statistics Tool or mentioning information that is only derivable from the Progress Statistics, for example the class average
	Concept trail	Explicitly mentioning the Concept Trail
	Other	All remaining comments
Specificity	Concepts	Mentioning concrete concepts from the discussion or the written text. Also includes quoting chat messages
	Comparison	Making a comparison between several groups, for example by comparing to the class average
	Development	Comments that make a reference to development within a group, which demonstrate that the participant diagnosed multiple times, or comments about sequences of actions
	Other	All remaining comments

Table 4
Average number of actions per participant per tool per vignette.

Tool	Selection/action	Control (*SD*)	Experimental (*SD*)
Chat tool	Opening chat	25.0 (14.0)	27.5 (9.8)
	Scrolling in chat	10.3 (13.6)	5.4 (5.3)
Coach menu	Content of cowriter	9.7 (4.5)	8.7 (4.2)
	Progress statistics cowriter	–	6.2 (3.7)
	Progress statistics chat	–	1.9 (1.6)

Participants could use the Coach menu to select the content of the groups' text editors and, in the experimental condition, the PS. The content of the cowriter was selected equally often in the two conditions ($t(38) = 0.71$, $p = .48$, $d = 0.23$). Within the experimental condition, participants selected one of the available statistics on average 8.1 times per vignette ($SD = 4.5$). The Progress Statistics of the Chat were consulted significantly less often than the Progress Statistics of the Cowriter; $t(19) = -5.61$, $p = .00$, $d = 1.51$.

3.2. Research question 1: diagnosis of concept coverage in the discussion and task progress

3.2.1. Overall diagnosis scores

At the end of each vignette, participants gave a score to each group for the concept coverage in the group's discussion and for the group's task progress. Each vignette contained one or more groups that showed deviating behavior concerning one of these aspects. Thus, we expected that the problematic group would receive a lower score than the other groups concerning the deviating aspect. We used a relative standard instead of an absolute standard for comparing the scores to take into account that different teachers might have different ideas of what constitutes a sufficient or an insufficient grade. In vignette 2 and 3, the problematic groups scored significantly lower than the non-problematic groups on both aspects in both conditions, so both on the deviating aspect as well as the non-problematic aspect ($p < .05$ in all cases). In vignette 1, only the problematic aspect (task progress) scored significantly lower in both conditions (control condition; $t(15) = -4.55$, $p < .001$, experimental condition; $t(17) = -2.73$, $p = .014$). It thus seems that participants in both conditions singled out the problematic groups and in most cases, judged these groups lower on task progress as well as on concept coverage in the discussion.

To see whether *individual* participants singled out the problematic group or not, a variable at participant level was created that indicates whether the problematic group received the lowest score from a particular participant in a particular vignette. Because there were four problematic groups, the maximum number of groups that could be detected was 4. On average, 1.9 groups ($SD = 1.0$) were detected in the control condition, compared to 2.1 in the experimental condition ($SD = 1.0$). This difference was not significant ($t(38) = -0.62$, $p = .54$, $d = 0.20$).

3.2.2. Focus of diagnosis comments

Along with each given score, participants could give a comment. These comments were coded for focus, specificity, and source. Table 5 displays the distribution of focus of comments accompanying the overall diagnosis scores, separated for the scores for concept coverage in the discussion (left) and task progress (right). Participants who did not give any comments (10 participants; 5 in each condition) were excluded from this analysis.

Concerning the discussion, all participants focused largely on whether or not students were on topic (*Discussion on/off task*). In the control condition, there is on average relatively less focus on the content of the discussion (*Discussion content*) than in the experimental condition (10.9% compared to 19.2%), although this difference is not significant ($t(27) = -1.14$, $p = .26$, $d = 0.44$). Furthermore, in the control condition there is significantly more focus on social activities when judging the discussion than in the experimental condition (30.1% compared to 12.3%); $t(27) = 2.08$, $p = .047$, $d = 0.79$).

In both conditions, task progress was relatively often judged both by considering the quality (*Text quality*, 30.6% and 33.4%) as well as the appearance (*Text appearance*, 46.8% and 45.1%) of the groups' written texts.

3.2.3. Specificity and source of diagnosis comments

Comments were also coded for their specificity. In the control condition, of 220 comments, 0.9% contained comparisons between groups, 6.4% contained mention of specific task-related concepts. A further 8.6% contained references to developments within groups, which means that the participants showed that they had diagnosed the situation in a group multiple times during a vignette. For example, one participant indicated that in vignette 1, for one of the groups "the discussion was on topic, although at the end the conversation was about something else". In the experimental condition, out of 259 comments, 7.3% contained comparisons between groups, 1.2% contained mention of specific task-related concepts and 7.3% contained references to developments within groups.

In the experimental condition, it was investigated whether the two LA tools were mentioned as the source for the participant's comment. The CT was mentioned in 1.6% of the comments, compared to 11.2% for the PS. For example, the CT was mentioned in vignette 3 in a comment about one of the groups: "They had a good, on topic discussion. That was also clear from the number of times they mentioned the concepts". An example of a comment including the PS was "This group was ahead of group 2, but their progress was less than class average". Because the PS show a class overview, it was in line with

Table 5
Focus of diagnosis comments about Task progress and Discussion.

Focus	Concept coverage in discussion				Task progress			
	Control condition (N = 14)		Experimental condition (N = 15)		Control condition (N = 12)		Experimental condition (N = 14)	
	M (%)	*SD* (%)	*M* (%)	*SD* (%)	*M* (%)	*SD* (%)	*M* (%)	*SD* (%)
Discussion content	10.9	13.7	19.2	23.0	1.2	2.6	0.0	0.0
Discussion on/off task	49.9	31.8	54.3	24.9	1.8	4.5	4.0	8.3
Text quality	0.7	2.3	0.8	2.2	30.6	30.6	33.4	22.8
Text appearance	0.0	0.0	0.5	1.9	46.8	29.7	45.1	30.6
Strategy	5.3	11.5	8.2	11.6	8.0	11.9	4.9	11.8
Social activities	30.1	28.7	12.3	13.7	8.4	19.6	6.0	8.7
Other	3.1	8.8	4.7	9.7	3.1	5.5	6.5	12.4

Note. Some participants only gave comments on either concept coverage or task progress. They were not excluded from the analysis, but it did mean that the N could differ for each column in this table.

expectations that out of 29 comments that included the PS, 9 of those included comparisons between groups. In contrast to our expectation, there were no comments that included both mention of the CT and mention of task-related concepts.

3.3. Research question 2: frequency, focus, and specificity of interventions

3.3.1. Frequency of interventions

The total amount of interventions is 523, of which 200 in the control condition and 323 in the experimental condition. Table 6 shows the mean number of interventions for each vignette. The mean frequency of interventions is significantly higher in the experimental condition ($t(38) = -2.07$, $p = .045$, $d = 0.67$).

In the control condition, the mean number of interventions sent to the problematic group does not significantly differ from the mean number of interventions sent to non-problematic groups ($t(16) = 1.049$, $p = .310$, $d = 0.24$). In the experimental condition, the problematic groups on average receive more interventions than the non-problematic groups overall ($t(19) = 5.041$, $p = .000$, $d = 0.33$). Comparing between conditions, the overall percentage of interventions aimed at the problematic group does not significantly differ.

3.3.2. Focus and specificity of interventions

Table 7 displays the distribution of focus of the interventions in both conditions. As can be seen, the percentages for the categories *Discussion content*, *Strategy*, and *Text quality* are quite similar. The largest differences between the two conditions can be found for *Discussion content* (16.0% compared to 5.4%) and *Text appearance* (18.8% compared to 9.4%). None of the mean percentages were significantly different ($p > .05$ in all cases).

Besides the focus on cognitive activities, there is a relatively high focus on social activities in both conditions (24.3% control condition, 20.4% experimental condition). These interventions can roughly be split into general compliments (41.1% and 38.6%), comments about collaborative processes such as division of tasks (43.5% and 47.1%) and motivating statements or offering of help (15.2% and 14.3%).

When comparing the subset of interventions concerning the quality of students' work (concept coverage in discussions and text quality, i.e., *Discussion content* and *Text quality*), it was found that in both conditions about one third of the interventions contains specific mention of concepts relating to the task. There is thus no significant difference in the average specificity of the interventions between the two conditions ($t(38) = 0.034$, $p = .97$, $d = 0.01$).

3.3.3. Interventions aimed at the problematic issue

For each problematic group, there were specific elements in the students' activities that could be noticed by the participants. It was coded whether participants indeed noticed and remarked upon these elements. For each intervention in a problematic group, it was coded whether it remarked upon the problematic issue of the group in terms of the combination of focus and means that the participant employed. Whether an intervention was aimed at the problem at hand depends on the

Table 6
Mean number of interventions per participant in each vignette.

Vignette	Mean number of interventions (*SD*)	
	Control condition	Experimental condition
1	2.6 (2.1)	4.5 (3.4)
2	3.6 (3.2)	5.8 (3.9)
3	4.3 (4.6)	5.9 (4.3)
Overall	3.3 (2.6)	5.4 (3.6)

Table 7
Distribution of focus of interventions in the two conditions.

Focus	Control (N = 18)		Experimental (N = 19)		*p*
	M (%)	*SD* (%)	*M* (%)	*SD* (%)	
Discussion content	5.4	9.6	16.0	22.1	0.070
Discussion on/off task	9.4	14.4	12.5	11.1	0.233
Text quality	25.9	25.3	20.1	14.9	0.641
Text appearance	18.8	26.2	9.4	9.2	0.075
Strategy	13.8	15.8	19.6	12.5	0.126
Social activities	24.3	26.2	20.4	14.4	0.988
Other	2.4	4.4	2.1	3.3	0.964

specific vignette and the specific group. For example, in vignette 1 there was a group that was behind concerning task progress, but in which the concept coverage in the discussion was very high. Thus, interventions to this group that were concerned with the problematic issue would include *instructions* or *suggestions* (means) concerning *Text appearance* (focus) as well as *positive evaluations* or *suggestions* (means) concerning *Discussion content* (focus).

Then, for each participant and each problematic group it was coded whether the participant did or did not remark upon the problematic issue by intervening. Because there were four problematic groups divided over three vignettes, the highest possible score was 4, and the lowest was 0. These scores were compared between the two conditions. In the control condition, the average score was 1.3 ($SD = 1.2$), compared to 2.0 ($SD = 1.2$) in the experimental condition. This difference was not significant; ($t(38) = -1.85$, $p = .072$, $d = 0.58$).

It was investigated whether the detection of a problematic groups would lead to an intervention concerning the problematic issue. It seems that especially in the control condition, giving the problematic group the lowest score did not mean that the participant would also intervene in the concerning group. In the control condition, when the group was detected, this co-occurred with an intervention in the concerning group in 29.7% of all cases, compared to 58.5% in the experimental condition.

3.4. Research question 3: cognitive load

On average, participants in the control condition reported a cognitive load of 6.4 ($SD = 1.8$) compared to 6.7 ($SD = 1.4$) in the experimental condition. There is no significant difference between the conditions ($t(38) = -0.65$, $p = .517$, $d = 0.19$). The average reported cognitive load for each of the separate vignettes lies between 6 and 7 in both conditions. Cognitive load did decrease with each successive vignette in the experimental condition, although this decrease between vignettes is not significant ($F(2, 38) = 1.39$, $p = .260$). Within the experimental condition, there was no relationship between the use of supporting tools and cognitive load ($r(58) = -0.13$, $p = .342$).

By means of logistic regression it was investigated whether cognitive load could contribute to predicting the detection of the problematic group, but this was not the case ($B = -0.13$ ($SE = 0.16$), $p = .936$). Furthermore, the mean cognitive load did not correlate with the mean number of detected problematic groups ($r(38) = -0.012$, $p = .941$) nor with how often participants remarked upon the problematic issue in problematic groups ($r(38) = -0.06$, $p = .712$).

4. Discussion

Teacher regulation of CSCL is a demanding task. Multiple collaborating groups perform multiple activities, which makes continuous diagnosis a task that could create high cognitive load. The present study examined whether teacher supporting tools could assist teachers by visualizing analyses of students' cognitive activities. The main research question was what the effects are of teacher supporting tools on teachers' diagnosis, interventions, and experienced cognitive load during CSCL.

4.1. Discussion of findings

Teachers were asked to give a diagnosis of each collaborating group for the concept coverage in the discussion and the general task progress. In each vignette, one or two groups had a problem concerning one of these two aspects. The results showed that providing teachers with supporting tools had no effect on their diagnosis of the groups. Participants in both conditions singled out the problematic groups. Thus, in contrast to expectation, it does not seem like the supporting tools increased the detection of problematic groups. Participants were able to give an explanation of their diagnosis scores, which were coded for their focus. No differences were found except for the category ‘social activities’, which meant that the participants with no access to the supporting tools more often focused on students' social activities when they judged students' discussions than the participants who had access to the tools did.

Concerning teacher interventions, we investigated frequency, focus and specificity. The mean frequency of interventions was significantly higher when teachers had access to the supporting tools, and those interventions were relatively more often directed at problematic groups than at non-problematic groups. There was also evidence for a further effect of the supporting tools, namely that when participants detected a problematic group, they were more likely to act upon this detection with an intervention focused at the problematic issue. Overall, no differences were found concerning the focus of interventions, neither were there any differences in the specificity of teacher interventions.

Our main hypothesis for the mechanism underlying these findings was that the supporting tools would lower cognitive load because they give an overview of students' cognitive activities that is easy to interpret. This, in turn, would lead to increased availability of mental effort. Cognitive load was measured at the end of each of the three vignettes as self-reports on a scale from 1 to 9. No significant differences were found between the teachers who had access to the supporting tools and those who did not. On average, cognitive load was perceived between 6 and 7 on a scale from 1 to 9, which means all participants indicated the task of regulating CSCL took a moderate amount of mental effort. Thus, in contrast to expectations, participants in both conditions reported equal amounts of mental effort. Multiple explanations can be offered for these findings.

First of all, it might be the case that the tools only lower cognitive load after longer periods of time or after a more extensive training in using the tools. None of the participants had much experience with CSCL, so besides getting used to the digital learning environment, effectively using the supporting tools might require time as well (for the effect of training on teaching in online settings, see De Smet, Van Keer, De Wever, & Valcke, 2010). It could also be argued that any dashboard or LA tool, despite being designed for reducing cognitive load, first increases the teacher's load because it increase the information displayed in the screen. However, there are some arguments against this explanation. First of all, participants were asked at the end of the experiment whether they understood the set-up and whether the practice vignette and the explanation at the beginning were clear. The participants indeed acknowledged that this was the case. Secondly, if initially the load was higher because of the LA tools, we would have expected to find a decline in reported cognitive load. This was not the case. This seems to indicate that the visual representation of the LA tools, together with the practice vignette, gave participants enough information to work with the LA tools. It could still be that the experiment as a whole was too short for other, more profound effects of the LA tools to appear. It is therefore important to investigate the tools in authentic settings, in which classroom lessons have a longer duration and student activities more clearly progress over time. Although we made use of authentic data to design the vignettes, it was the case that students could not respond to the teachers' interventions. There are no indications that this posed an obstacle to the teachers. On the contrary, participants continued to send interventions after the first vignette, and often sent multiple interventions per vignette, leading us to believe that the situation was not awkward to an extent that this would prevent participants from intervening. However, the fact remains that this study was a controlled experiment, and that further studies in classrooms are necessary.

Another interpretation of the results concerning cognitive load could be that the cognitive load *was* lowered. After all, in the experimental condition, participants intervened more. Diagnosing and intervening both cost mental effort. So although the invested mental effort was the same, the way the effort was spent may have been different. In the control condition all effort went to diagnosing, whereas in the experimental condition there was effort left to spend on interventions. The measurement of cognitive load in this study in the form of a single score based on self-reports, which is a common method to measure cognitive load (Van Gog & Paas, 2012), does not enable us to confirm one of these explanations. We could have specified two questions so that participants would indicate the effort associated with diagnosing and intervening separately.

4.2. Conclusions and implications

Together, these results seem to imply that although the teacher supporting tools under study, the Concept Trail and the Progress Statistics, did not improve the ability to detect relevant events, they *did* stimulate teachers to act: the teachers in the experimental condition intervened significantly more often. Of course, the question is whether more teacher interventions are beneficial for the collaboration between students. Based on the present findings, we cannot univocally answer this question. We did find that teachers in the experimental condition more often intervened in problematic groups than in non-problematic groups, which could indicate they were better able to direct their attention at groups that needed it most. As simultaneous diagnosis of multiple groups is a demanding task (Schwarz & Asterhan, 2011), this is an important finding. When a problem was detected, the teachers in the experimental condition twice as often sent a message to the group concerning this problem than in the control condition. These results thus seem to point at least to the conclusion that the supporting tools led teachers not only to detect, but also to address problematic issues. Whether this was a support to the collaborating groups remains the question. Even a few well-chosen interventions may have beneficial effects (Lin et al., 2015). As said, because we used simulations of collaboration, we could not investigate the students' response on teacher interventions.

As discussed in Section 4.1, if lowered cognitive load did not account for the effects of the supporting tools on the frequency of interventions, two other hypotheses can be offered. First of all, it might be the case that teachers with the learning analytics tools felt more confident about their diagnosis. After all, they had more ‘evidence’ to support their thoughts and more specific information to act on. Thus, there may have been enough information about the collaborative process for the teachers in the control condition to detect the problematic groups, but the supporting tools provided the teachers in the experimental condition with additional confidence to intervene. If this was the case, we would also have expected the specificity of the interventions in the experimental condition to increase, which was not found. Teachers could for example have used the concepts from the Concept Trail to provide students with more detailed feedback, which increases the intervention's effectiveness (Voerman et al., 2012).

Another explanation could be that besides summarizing students' cognitive activities, the mere presence of the supporting tools steered the participants' focus more towards cognitive aspects of collaboration, which led them to intervene more on the occurrence of problems concerning cognitive activities. This was evidenced by the finding that participants without the supporting tools focused more on social activities in their diagnosis of the groups' discussions. It is important that teachers pay attention to both cognitive and social activities, especially in a collaborative setting (Asterhan, Schwarz, & Gil, 2012). In this study, we explicitly wanted to examine the effect of tools showing visualizations of cognitive activities. Here we see the prescriptive function that learning analytics tools can serve (Duval, 2011): by choosing a particular tool as a researcher, implicitly this means there is an idea about what information is important. In class situations, tools could be offered to teachers concerning both cognitive and social activities, with the additional option for teachers to choose which tools are displayed or not, depending on the needs of the situation.

This study has provided first steps into investigating the effects of learning analytics on teacher regulation. Of course, there are still important directions for future research to consider that cannot be answered based on this controlled experimental study alone. It could be argued that context factors such as the size of the classroom, the type of learning activity and the type of collaboration students engage in, and the kind of learning analytics all influence how the teacher regulates student activity. For example, increased class size with more than five groups of students heightens the need for additional teacher support (Van Leeuwen, Janssen, Erkens, & Brekelmans, 2015; Chounta & Avouris, 2014). Also, collaboration in synchronous versus asynchronous or face-to face versus online settings may change the demands on teachers and subsequently how teachers may best be supported (Solimeno, Mebane, Tomai, & Francescato, 2008). For now, the possible explanations for how teachers use learning analytics discussed in the paragraphs above do provide a theoretical framework to think about the relationship between what a teacher does and how learning analytics may support this process.

4.3. Limitations and directions for future research

Our study has some methodological limitations, which could be taken into account in future research. The first limitation concerns the study's sample, which consisted of undergraduate educational sciences students with limited teaching experience. The question is thus whether other results would have been obtained if experienced teachers had participated in this study. There are some studies that have investigated the role of experience in online teaching, but the results are not univocal. From these studies, it appeared for example that novice teachers experience more challenges and felt uncertain about their skills as regulator (De Laat, 2006; De Smet et al., 2010). Also, novice and experienced teachers differed in the importance they ascribed to the different types of student learning activities (Kopp, Matteucci, & Tomasetto, 2012), and differed in the average number of interventions they perform, with novice teachers intervening less than experienced ones (Goold, Coldwell, & Craig, 2010). However, in an earlier study with a similar setup, we found no differences between beginning student teachers and more experienced teachers concerning diagnosing and intervening strategies (Van Leeuwen et al., 2014). The role of teaching experience is therefore an important direction for future research. Regarding the present study, we have made sure that the two conditions did not differ concerning teaching experience, so that the results could not be ascribed to this factor.

The second limitation concerns our measurement of teachers' diagnosis. While teacher interventions during CSCL are always observable, the diagnoses preceding interventions are not, because it is a mental process that is usually not explicated. In this study, we asked teachers about their diagnosis after each vignette, so they gave an overall diagnosis, not a real time one. This measure of 'detection' was rather crude and only consisted of two overall scores with short additional explanations. When participants scored a group low, they usually did so for both the problematic aspect as well as the non-problematic aspect. Although the vignettes were chosen so that the problematic group only showed deviating behavior on either concept coverage or task progress, in most cases the group that was judged lowest by the participants received low scores on *both* aspects. This might be a sign that the overall diagnosis was constructed after the vignettes instead of being a reflection of the actual, real-time diagnosis. An alternative explanation is the halo-effect, which means that a prior negative evaluation of a group of students biases the subsequent grading of the same students on a different aspect or task (Malouff, Emmerton, & Schutte, 2013). A challenge for future research is to come up with a method to also measure moment-to-moment diagnoses. A possibility would be to use stimulated interviews or thinking aloud methods (Schwarz & Asterhan, 2011). This would increase the possibility of adequately researching whether diagnoses becomes more specific with the addition of learning analytics tools. It would also give insight into how teachers integrate the knowledge from the LA tools with the knowledge available through the other tools within VCRI. It is known that for students, this process of combining information from different sources requires the integration of multiple representations into a coherent mental representation (Bodemer, Ploetzner, Feuerlein, & Spada, 2004). Thinking aloud methods could offer more details about what this process looks like for teachers.

As was said, the same critique of a lack of real-time measures applies to the way we measured cognitive load, because it was also measured as an overall score instead of a real-time one. One of our interpretations of the results was that teachers with access to LA tools may have felt more confident about their diagnoses. Therefore, instead of only asking about how difficult the task was, it might be more appropriate to also ask how confident teachers were about their choices. The difficulty of assessing both the cognitive load as well as the diagnosis itself in a real-time fashion is that interrupting the teacher to measure these constructs disrupts the completion of the task, which could hinder the validity of the results. For cognitive load, a solution could be to make use of physiological indicators that can unobtrusively be measured (such as stress levels) in addition to self-reports (Haapalainen, Kim, Forlizzi, & Dey, 2010).

The present experimental setup was used in a previous study (Van Leeuwen et al., 2014) to examine the effects of teacher supporting tools that show visualizations of students' *social* activities. Some of the expected effects that we did not currently find, *were* present in this previous study, for example that interventions were more specific in nature and that participants were better able to detect problems within groups when they had access to the supporting tools. Of course, the vignettes and the types of problems that the groups of students experienced were different, which makes it hard to draw clear comparisons between the two studies. A result that was found in both studies was that the relative amount of interventions aimed at problematic groups was higher than those aimed at the non-problematic groups when the participants had access to the supporting tools. This seems to indicate that at the level of moment-to-moment diagnosis and intervention, the investigated tools do in some way help teachers to detect occurrence of problems and that they do influence teachers' choice to intervene. With additional measures described in the paragraphs above, we could study whether the tools help to make the diagnosis easier, or whether the tools have more effect *after* diagnosing in terms of giving teachers more confidence to act. In general, the employed experimental methodology has proven valuable because the use of vignettes makes it possible to research the effects of teacher supporting tools in a systematic way. An interesting next step would be to study the effects of the LA tools in a real CSCL setting, in which students are able to respond to teacher interventions and in which group processes change and unfold over time (Chan, 2011). This could also provide more information concerning the question whether an increased number of teacher interventions has a positive effect on student collaboration.

4.4. Concluding remarks

To conclude, in the present study several effects of two learning analytics tools were found on teacher regulation of CSCL. After diagnosis of problems concerning students' cognitive activities, teachers with access to these tools were more likely to offer assistance to the students concerning these problems. Two explanations for this finding were discussed, namely that the learning analytics steered the teachers' focus towards cognitive activities, and that the tools increased the teachers' confidence of their diagnosis. Using an experimental set-up, we have therefore made first steps to uncovering the underlying mechanisms of how teachers are supported by learning analytics. In future studies hopefully these results can be expanded to include other types of tools and settings and eventually, be applied in authentic educational settings.

References

Anderson, T., Rourke, L., Garrison, D. R., & Archer, W. (2001). Assessing teaching presence in a computer conferencing context. *Journal of Asynchronous Learning Networks, 5*(2), 1–17.

Asterhan, C. S. C., & Eisenmann, T. (2011). Introducing synchronous e-discussion tools in co-located classrooms: a study on the experiences of “active” and “silent” secondary school students. *Computers in Human Behavior, 27*(6), 2169–2177. http://dx.doi.org/10.1016/j.chb.2011.06.011.

Asterhan, C. S. C., & Schwarz, B. B. (2010). Online moderation of synchronous e-argumentation. *International Journal of Computer-Supported Collaborative Learning, 5*, 259–282. http://dx.doi.org/10.1007/s11412-010-9088-2.

Asterhan, C. S. C., Schwarz, B. B., & Gil, J. (2012). Small-group, computer-mediated argumentation in middle-school classrooms: the effects of gender and different types of online teacher guidance. *British Journal of Educational Psychology, 82*(3), 375–397. http://dx.doi.org/10.1111/j.2044-8279.2011.02030.x.

Baker, M., Andriessen, J., Lund, K., Van Amelsvoort, M., & Quignard, M. (2007). Rainbow: a framework for analyzing computer-mediated pedagogical debates. *International Journal of Computer-Supported Collaborative Learning, 2*, 315–357. http://dx.doi.org/10.1007/s11412-007-9022-4.

Baker, R. S., Corbett, A. T., Koedinger, K. R., & Wagner, A. Z. (2004). Off-task behavior in the cognitive tutor classroom: when students “Game the system”. In *Proceedings of the Conference for Computer-Human interaction, April 24–29, Vienna, Austria* (pp. 383–390). http://dx.doi.org/10.1145/985692.985741.

Bodemer, D., Ploetzner, R., Feuerlein, I., & Spada, H. (2004). The active integration of information during learning with dynamic and interactive visualisations. *Learning and Instruction, 14*, 325–341. http://dx.doi.org/10.1016/j.learninstruc.2004.06.006.

Buder, J., & Bodemer, D. (2008). Supporting controversial CSCL discussions with augmented group awareness tools. *International Journal of Computer-Supported Collaborative Learning, 3*, 123–139. http://dx.doi.org/10.1007/s11412-008-9037-5.

Carroll, J. (1963). A model for school learning. *Teachers College Record, 64*, 723–733.

Casamayor, A., Amandi, A., & Campo, M. (2009). Intelligent assistance for teachers in collaborative e-learning environments. *Computers & Education, 53*, 1147–1154. http://dx.doi.org/10.1016/j.compedu.2009.05.025.

Chan, C. K. K. (2011). Bridging research and practice: implementing and sustaining knowledge building in Hong Kong classrooms. *International Journal of Computer-Supported Collaborative Learning, 6*, 147–186. http://dx.doi.org/10.1007/s11412-011-9121-0.

Chatti, M. A., Dyckhoff, A. L., Schroeder, U., & Thüs, H. (2012). A reference model for learning analytics. *International Journal of Technology Enhanced Learning, 4*, 318–331. http://dx.doi.org/10.1504/IJTEL.2012.051815.

Chounta, I.-A., & Avouris, N. (2014). *Towards the real-time evaluation of collaborative activities: Integration of an automatic rater of collaboration quality in the classroom from the teacher's perspective*. Education and Information Technologies. http://dx.doi.org/10.1007/s10639-014-9355-3. Online preprint.

Cohen, E. G. (1994). Restructuring the classroom: conditions for productive small groups. *Review of Educational Research, 64*(1), 1–35. http://dx.doi.org/10.3102/00346543064001001.

Coll, C., Rochera, M. J., & de Gispert, I. (2014). Supporting online collaborative learning in small groups: teacher feedback on learning content, academic task and social participation. *Computers & Education, 75*, 53–64. http://dx.doi.org/10.1016/j.compedu.2014.01.015.

De Groot, R., Drachman, R., Hever, R., Schwarz, B., Hoppe, U., Harrer, A., et al. (2007). Computer supported moderation of E-discussions: the ARGUNAUT approach. In *Proceedings of the Conference on Computer-Supported Collaborative Learning (CSCL 2007)*.

De Laat, M. (2006). *Networked learning*. Apeldoorn: Politie Academy.

De Smet, M., Van Keer, H., De Wever, B., & Valcke, M. (2010). Cross-age peer tutors in asynchronous discussion groups: exploring the impact of three types of tutor training on patterns in tutor support and on tutor characteristics. *Computers & Education, 54*(4), 1167–1181. http://dx.doi.org/10.1016/j.compedu.2009.11.002.

De Wever, B., Schellens, T., Valcke, M., & Van Keer, H. (2006). Content analysis schemes to analyze transcripts of online asynchronous discussion groups: a review. *Computers & Education, 46*(1), 6–28. http://dx.doi.org/10.1016/j.compedu.2005.04.005.

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. http://dx.doi.org/10.1016/j.chb.2010.05.018.

Doyle, W. (2006). Ecological approaches to classroom management. In C. M. Evertson, & C. S. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 97–125). Mahwah, NJ: Erlbaum.

Duval, E. (2011). Attention Please! Learning analytics for visualization and recommendation. In *Proceedings of the 1st International Conference on Learning Analytics and Knowledge* (pp. 9–17). NY, USA: ACM New York. http://dx.doi.org/10.1145/2090116.2090118.

Dyckhoff, A. L., Zielke, D., Bültmann, M., Chatti, M. A., & Schroeder, U. (2012). Design and implementation of a learning analytics toolkit for teachers. *Educational Technology & Society, 15*(3), 58–76.

Erkens, G., Jaspers, J., Prangsma, M., & Kanselaar, G. (2005). Coordination processes in computer supported collaborative writing. *Computers in Human Behavior, 21*, 463–486. http://dx.doi.org/10.1016/j.chb.2004.10.038.
Feldon, D. F. (2007). Cognitive load and classroom teaching: the double-edged sword of automaticity. *Educational Psychologist, 42*(3), 123–137. http://dx.doi.org/10.1080/00461520701416173.
Gibbs, G., & Simpson, C. (2004). Conditions under which assessment supports students' learning. *Learning and Teaching in Higher Education, 1*, 3–31.
Goold, A., Coldwell, J., & Craig, A. (2010). An examination of the role of the e-tutor. *Australasian Journal of Educational Technology, 26*(5), 704–716. Retrieved from http://www.ascilite.org.au/ajet/ajet26/goold.html.
Haapalainen, E., Kim, S., Forlizzi, J. F., & Dey, A. K. (2010). Psycho-physiological measures for assessing cognitive load. In *Proceedings of the 12th ACM International Conference on Ubiquitous Computing* (pp. 301–310). Copenhagen: ACM. http://dx.doi.org/10.1145/1864349.1864395.
Harrer, A., McLaren, B. M., Walker, E., Bollen, L., & Sewall, J. (2006). Creating cognitive tutors for collaborative learning: steps toward realization. User modeling and user-adapted interaction. *The Journal of Personalization Research, 16*, 175–209.
Hsieh, Y., & Tsai, C. (2012). The effect of moderator's facilitative strategies on online synchronous discussions. *Computers in Human Behavior, 28*(5), 1708–1716. http://dx.doi.org/10.1016/j.chb.2012.04.010.
Janssen, J., Erkens, G., Kanselaar, G., & Jaspers, J. (2007). Visualization of participation: does it contribute to successful computer-supported collaborative learning? *Computers & Education, 49*(4), 1037–1065. http://dx.doi.org/10.1016/j.compedu.2006.01.004.
Jermann, P., Soller, A., & Muehlenbrock, M. (2005). From mirroring to guiding: a review of state of the art technology for supporting collaborative learning. *International Journal of Artificial Intelligence in Education, 15*(4), 261–290. Retrieved from http://iospress.metapress.com/content/dyrlcql4wxwutcqc/.
Kaendler, C., Wiedmann, M., Rummel, N., & Spada, H. (2014). Teacher competencies for the implementation of collaborative learning in the classroom: a framework and research review. *Educational Psychology Review*. http://dx.doi.org/10.1007/s10648-014-9288-9 (online preprint).
Kirschner, P. A., & Erkens, G. (2013). Toward a framework for CSCL research. *Educational Psychologist, 48*(1), 1–8. http://dx.doi.org/10.1080/00461520.2012.750227.
Kopp, B., Matteucci, M. C., & Tomasetto, C. (2012). E-tutorial support for collaborative online learning: an explorative study on experienced and inexperienced e-tutors. *Computers & Education, 58*(1), 12–20. http://dx.doi.org/10.1016/j.compedu.2011.08.019.
Lin, T.-J., Jadallah, M., Anderson, R. C., Baker, A. R., Nguyen-Jahile, K., Kim, I.-H., et al. (2015). Less is more: teachers' influence during peer collaboration. *Journal of Educational Psychology, 107*(2), 609–629. http://dx.doi.org/10.1037/a0037758.
Malouff, J. M., Emmerton, A. J., & Schutte, N. S. (2013). The Risk of a Halo Bias as a reason to keep students anonymous during grading. *Teaching of Psychology, 40*(3), 233–237. http://dx.doi.org/10.1177/0098628313487425.
Mazza, R., & Dimitrova, V. (2007). CourseVis: a graphical student monitoring tool for supporting instructors in web-based distance courses. *International Journal of Human-Computer Studies, 65*(2), 125–139. http://dx.doi.org/10.1016/j.ijhcs.2006.08.008.
Michinov, N., & Primois, C. (2005). Improving productivity and creativity in online groups through social comparison process: new evidence for asynchronous electronic brainstorming. *Computers in Human Behavior, 21*(1), 11–28. http://dx.doi.org/10.1016/j.chb.2004.02.004.
Onrubia, J., & Engel, A. (2009). Strategies for collaborative writing and phases of knowledge construction in CSCL environments. *Computers & Education, 53*, 1256–1265. http://dx.doi.org/10.1016/j.compedu.2009.06.008.
Onrubia, J., & Engel, A. (2012). The role of teacher assistance on the effects of a macro-script in collaborative writing tasks. *International Journal of Computer-Supported Collaborative Learning, 7*(1), 161–186. http://dx.doi.org/10.1007/s11412-011-9125-9.
Paas, F., Tuovinen, J., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*, 63–72. http://dx.doi.org/10.1207/S15326985EP3801_8.
Papamitsiou, Z., & Economides, A. A. (2014). Learning analytics and educational data mining in practice: a systematic literature review of empirical evidence. *Educational Technology & Society, 17*(4), 49–64.
Pargman, T. C. (2003). Collaborating with writing tools: an instrumental perspective on the problem of computer-supported collaborative activities. *Interacting with Computers, 15*(6), 737–757. http://dx.doi.org/10.1016/j.intcom.2003.09.003.
Petrou, A., & Dimitracopoulou, A. (2003). Is synchronous computer mediated collaborative problem-solving 'Justified' only when by distance? Teachers' points of views and interventions with CO-located groups, during everyday class activities. In *Proceedings of the International Conference on Computer Support for Collaborative Learning, 2003* (pp. 441–450). Netherlands: Springer.
Romero, C., & Ventura, S. (2010). Educational data mining: a review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, 40*(6), 601–618. http://dx.doi.org/10.1109/TSMCC.2010.2053532.
Rummel, N., & Spada, H. (2005). Learning to collaborate: an instructional approach to promoting collaborative problem solving in computer-mediated settings. *Journal of the Learning Sciences, 14*(2), 201–241. http://dx.doi.org/10.1207/s15327809jls1402.
Salinas, M. F. (2008). From Dewey to Gates: a model to integrate psychoeducational principles in the selection and use of instructional technology. *Computers & Education, 50*, 652–660. http://dx.doi.org/10.1016/j.compedu.2006.08.002.
Schwarz, B. B., & Asterhan, C. A. (2011). E-Moderation of synchronous discussions in educational settings: a Nascent practice. *Journal of the Learning Sciences, 20*(3), 395–442. http://dx.doi.org/10.1080/10508406.2011.553257.
Siemens, G., & Gasevic, D. (2012). Guest editorial – learning and knowledge analytics. *Educational Technology & Society, 15*(3), 1–2. Retrieved from http://www.ifets.info/journals/15_3/1.pdf.
Solimeno, A., Mebane, M. E., Tomai, M., & Francescato, D. (2008). The influence of students and teachers characteristics on the efficacy of face-to-face and computer supported collaborative learning. *Computers & Education, 51*(1), 109–128. http://dx.doi.org/10.1016/j.compedu.2007.04.003.
Stahl, G., Koschmann, T., & Suthers, D. (2006). Computer-supported collaborative learning: an historical perspective. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (pp. 409–426). Cambridge, UK: Cambridge University Press.
Van Gog, T., & Paas, F. (2012). Cognitive load measurement. In N. Seel (Ed.), *Encyclopedia of the sciences of learning* (pp. 599–601). New York: Springer.
Van Leeuwen, A., Janssen, J., Erkens, G., & Brekelmans, M. (2013). Teacher interventions in a synchronous, co-located CSCL setting: Analyzing focus, means, and temporality. *Computers in Human Behavior, 29*(4), 1377–1386. http://dx.doi.org/10.1016/j.chb.2013.01.028.
Van Leeuwen, A., Janssen, J., Erkens, G., & Brekelmans, M. (2014). Supporting teachers in guiding collaborating students: Effects of learning analytics in CSCL. *Computers & Education, 79*, 28–39. http://dx.doi.org/10.1016/j.compedu.2014.07.007.
Van Leeuwen, A., Janssen, J., Erkens, G., & Brekelmans, M. (2015). Teacher regulation of multiple computer-supported collaborating groups. *Computers in Human Behavior, 52*, 233–242. http://dx.doi.org/10.1016/j.chb.2015.05.058.
Verbert, K., Govaerts, S., Duval, E., Santos, J. L., Van Assche, F., Parra, G., et al. (2014). Learning dashboards: an overview and future research opportunities. *Personal and Ubiquitous Computing, 18*, 1499–1514. http://dx.doi.org/10.1007/s00779-013-0751-2.
Vermunt, J., & Verloop, N. (1999). Congruence and friction between learning and teaching. *Learning and Instruction, 9*(3), 257–280. http://dx.doi.org/10.1016/S0959-4752(98)00028-0.
Vermunt, J. D., & Vermetten, Y. J. (2004). Patterns in student learning: relationships between learning strategies, conceptions of learning, and learning orientations. *Educational Psychology Review, 16*(4), 359–384. http://dx.doi.org/10.1040-726X/04/1200-0359/0.
Voerman, L., Meijer, P. C., Korthagen, F. A. J., & Simons, P. R. J. (2012). Types and frequencies of feedback interventions in classroom interaction in secondary education. *Teaching and Teacher Education, 28*(8), 1107–1115. http://dx.doi.org/10.1016/j.tate.2012.06.006.
Voyiatzaki, E., & Avouris, N. (2014). Support for the teacher in technology-enhanced collaborative classroom. *Education and Information Technologies, 19*(1), 129–154.
Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education, 46*(1), 71–95. http://dx.doi.org/10.1016/j.compedu.2005.04.003.