International Journal of Computer-Supported Collaborative Learning
https://doi.org/10.1007/s11412-019-09299-x

What information should CSCL teacher dashboards provide to help teachers interpret CSCL situations?

Anouschka van Leeuwen[1] • Nikol Rummel[2] • Tamara van Gog[1]

Received: 29 July 2018 / Accepted: 6 May 2019

Abstract
Teachers play a major role during CSCL by monitoring and stimulating the types of interactions between students that are conducive to learning. Teacher dashboards are increasingly being developed to aid teachers in monitoring students' collaborative activities, thereby constituting a form of indirect support for CSCL in the classroom. However, the process of how teachers find and interpret relevant information from dashboards, and which help they need during this process, remains largely unexamined. We first describe how we arrived at the design of a prototype teacher dashboard in the context of primary school fraction assignments, based on teacher interviews. We then report an experimental study ($n = 53$) to investigate the effect of the type of support a teacher dashboard offers (mirroring, alerting, or advising) on teachers' detection and interpretation of potentially problematic situations where intervention might be needed. The results showed that the type of support did not influence detection of events, but did influence teachers' interpretation of a CSCL situation.

Keywords Collaborative learning · Teacher support · Learning analytics · Dashboard

Introduction

Collaborative learning is an instructional strategy shown to have positive effects on student achievement (Kyndt et al. 2014). By collaboratively completing a task, students are challenged to share ideas, express their thoughts and engage in discussion, which contributes to learning. Using *computer-supported* collaborative learning (CSCL) can help to stimulate these processes because the technology can support shared activities of exploration and social interaction (Stahl et al. 2006). but requires adequate support to lead to the development of the intended knowledge and skills (Gillies et al. 2008). Teachers play a major role during CSCL by

✉ Anouschka van Leeuwen
A.vanLeeuwen@uu.nl

[1] Department of Education, Utrecht University, Heidelberglaan 1, 3584CS, Utrecht, the Netherlands

[2] Pädagogische Psychologie, Institut für Erziehungswissenschaft, Ruhr-Universität Bochum, Universitätsstraße 150, D - 44801, Bochum, Germany

monitoring and stimulating the types of interactions between students that are conducive to learning (Gillies et al. 2008). Teachers need to monitor the interaction between students so that they can carefully calibrate their pedagogical strategies to each group of students (Kaendler et al. 2015). It is therefore essential that teachers find relevant information quickly and accurately and derive the right inferences about the needs of their students. Most CSCL environments allow tracking of student behavior, which can be analyzed and fed back to the teacher in so-called *teacher dashboards* to inform them of the activities of collaborating students. Teacher dashboards are visual displays that provide teachers with information about their students to aid them with monitoring their students' progress (Verbert et al. 2014). Teacher dashboards can thereby be regarded as technological artifacts that *indirectly* support CSCL (Rummel 2018): by informing the teacher, they enhance the teacher's cognitive representation of the situation, and create the necessary preconditions for enabling the teacher to better attend to the needs of the collaborating students. Investigating how teachers interpret information from teacher dashboards and subsequently use it to inform their pedagogical actions, can thus contribute to the overarching goal of successfully implementing CSCL in the classroom.

However, the process of how teachers find and interpret relevant information about collaborating students on these dashboards remains largely unexamined (Van Leeuwen et al. 2017). It is essential teachers find relevant information quickly and accurately and derive the right inferences about their students. In terms of the process of *teacher noticing* (Van Es and Sherin 2008), the question is how teachers detect and interpret information on the dashboard and to what extent they need help doing so. The goal of the present paper is to contribute to our knowledge about these issues by investigating the effect of three types of interpretational aids on teacher noticing processes of collaborative situations. We thus aim to contribute to knowledge about the role of the teacher in such situations and, thereby, to possible implementation of CSCL support through support of the teacher.

The present study is carried out in the context of primary school student collaboration on fraction assignments in a CSCL setting. The teacher dashboard prototype we developed provides information about various aspects of the activities of the collaborating students (e.g., number of attempts on a task and amount of talk within a group), thereby offering insight into cognitive and social aspects of collaboration. We systematically study how teachers interpret initial versions of a teacher dashboard that display simulated, fictitious situations. This type of controlled experiment allows us to gain more in-depth insight into how teachers make sense of CSCL situations, which can inform the next steps in implementing teacher dashboards that enable more effective teacher support of CSCL. After discussing the theoretical underpinnings of our study in more depth, we describe the development of the dashboard prototype from which the three versions of interpretational aid were created for the experimental study.

Theoretical background

Teacher support of CSCL and the role of CSCL teacher dashboards

Teachers play an essential role during CSCL (Gillies et al. 2008), and this role can be broken down into three phases, each requiring specific teacher competencies (Kaendler et al. 2015). Prior to collaboration, teachers need to prepare, for example, the task materials and plan student grouping. In the interactive phase - during student collaboration - teachers need to monitor and support collaborative activities. Finally, after the interactive phase is terminated, teachers need to reflect on the effectiveness of students' collaboration and on their own role, which serves as input for future preparation phases. Focusing on the interactive phase, one of the core competencies required of teachers (Kaendler et al. 2015) is to *monitor* the collaborating students in order to make *informed decisions* about supporting students. A number of studies have closely examined the complexity of decisions and considerations teachers face whilst guiding CSCL. Greiffenhagen (2012) describes how teachers make rounds through the classroom, engaging in short or longer durations of teacher-student interaction, using these interactions as well as non-verbal behavior and observation of student work to constantly monitor what is happening. Van Leeuwen et al. (2015a) describe how teachers use information about their students to both proactively initiate teacher-student interactions as well as to reactively respond to students' request for support. There seems to be consensus arising from research that to inform teachers' actions, it is essential for teachers to stay up to date with each collaborating group's activity, which can be broken down into students' cognitive, metacognitive, and social activities (Kaendler et al. 2015). Simultaneously, there is also consensus about the demanding nature of guiding collaborating students, as teachers' time and cognitive resources are limited (Feldon 2007; Gillies and Boyle 2010; Greiffenhagen 2012; Van Leeuwen et al. 2015a).

Most CSCL environments allow tracking of student behavior, which can be analyzed and fed back to the teacher to support monitoring. By aggregating, analyzing, and displaying information about students' collaborative activities that are collected through the digital traces students leave behind, teachers' understanding of those activities may be enhanced. The idea of collecting information about learners to inform teachers stems from a larger body of research in the field of learning analytics, which concerns itself with the analysis of digital traces to optimize learning and the environment in which it occurs (Siemens and Gašević 2012). Learning analytics is a broad, interdisciplinary field in which the general aim is to better understand and improve learning processes through data-driven insights (Lang et al. 2017). One application of learning analytics is the development of *dashboards*, which are visual interfaces that "capture and visualize traces of learning activities, in order to promote awareness, reflection, and sense-making" (Verbert et al. 2014, p. 1499). The agent that is supported by the dashboard may differ; while many studies have aimed at providing students with dashboards that allow them to monitor and regulate their own learning, there is a movement towards developing teacher dashboards as well (Tissenbaum et al. 2016; Wise and Vytasek 2017).

In this article, we define *CSCL teacher dashboards* as visual displays that provide teachers with information about their collaborating students to aid teachers in monitoring their students' progress in CSCL settings. Specifically during the interactive phase of CSCL, the underlying idea of teacher dashboards is to offer an overview of the activities of collaborating students, and to increase the accuracy and depth of the teacher's interpretation of the situation (Van Leeuwen 2015). In turn, this is expected to help teachers to timely intervene in groups who might need support. In terms of the typology by Rummel (2018), teacher dashboards can thereby be regarded as technological artifacts that indirectly support student collaboration: by informing the teacher, conditions are created that enable the teacher to support CSCL more effectively.

Teacher noticing in the context of CSCL teacher dashboards

When administering a CSCL activity, it is important that teachers continuously monitor their students' activity to provide effective support. They may consult a teacher dashboard in real time to obtain an impression of the status and progress of the collaborating students. The underlying assumption of teacher dashboards as collaborative support is that teachers' representation of the situation is influenced and enhanced by information shown on the dashboard. CSCL teacher dashboards therefore hold the promise of providing 'vision' to teachers (Duval 2011; Van Leeuwen et al. 2017) that would allow them to ultimately better attend to the needs of their students, for example by providing additional explanation about the task material (cognitive), by aiding students to discuss strategies for solving the task (metacognitive), or by stimulating students to discuss their answers with each other (social). With more information about students' activities at the teacher's disposal during the collaborative activity, the teacher is expected to be better able to select the most appropriate type of support at a given moment (i.e., initiating adaptive instructional interventions; Matuk et al. 2015). It is therefore essential that teachers are able to interpret information about students on CSCL teacher dashboards, which is the process we zoom in on in this article.

The *teacher noticing framework* describes the process and characteristics of teachers' interpretation of a classroom situation (Van Es and Sherin 2002, 2008). In this framework, a distinction is made between *detection*, deciding what is noteworthy and deserves further attention, and *interpretation*, reasoning about events and making connections between specific events and the broader principles they represent. Several characteristics can be identified to describe teachers' analyses of a situation (Van Es and Sherin 2008). Concerning detection, which is primarily action-oriented, teachers may focus on different students in the classroom and on different aspects of student behavior. Concerning interpretation, which is primarily knowledge-oriented, interpretations can be more or less specific and make use of more or less evidence from the observed situation. Teachers may also adopt a different stance to analyze a situation, ranging from describing to evaluating or interpreting events. When teachers take an interpretative stance, they connect situations to principles of teaching and learning rather than regarding them as isolated events (Van Es and Sherin 2002). Teachers thereby delve deeper into students' understanding of the subject matter and how that understanding came about, which is regarded to be beneficial for the effectiveness of subsequent support the teacher offers to students (Putnam and Borko 2000; Van Es and Sherin 2002). As Van Es and Sherin (2008) describe, because taking an interpretative stance is not what teachers automatically do or are always capable of, many studies have aimed at enhancing teachers' noticing so that teachers adapt their instruction to what is happening in the classroom.

The process of teacher noticing can be applied both to the situation in which the teacher observes student behavior in the classroom during CSCL, as well as to the situation in which the teacher consults a teacher dashboard to be informed of students' behavior. In the latter case, similar to the process of teacher noticing, the teacher needs to detect and interpret the information on the dashboard in such a way to arrive at a decision on whether, and if so, what type of support groups might need. When using the dashboard as a source of information, teachers need to find relevant information quickly and derive the right inferences. Otherwise, instead of a mechanism of aid, the dashboard can become an "obstacle" or a source of additional workload when interpreting information is not a fluent, easy process (Hoogland et al. 2016). The process of how teachers find and interpret relevant information on CSCL teacher dashboards remains largely unexamined (Van Leeuwen et al. 2017). It is therefore important to investigate the process of teacher noticing in the context of teacher dashboards and how the design of the dashboard may make the process of interpretation easy or easier for the teacher.

CSCL teacher dashboards and different levels of interpretational aid

There are different levels of aid that a dashboard can provide for the process of noticing (Van Leeuwen and Rummel in press; Verbert et al. 2013, 2014). To explain these levels of aid, we first discuss the basic algorithm underlying technological assistance tools for collaborative learning (in this case CSCL teacher dashboards) as described by Soller et al. (2005). The first step is that data about student collaboration are collected, which are analyzed and in some form displayed. Next, the data are compared to the desired state of student collaboration to detect certain events. When any deviations from the desired state occur, the deviations need to be interpreted and subsequently, a decision can be made whether any action is needed to support student collaboration. There is thus a parallel to the steps in the process of teacher noticing described above: information about students is made available to the teacher, and subsequently, relevant information or events need to be detected, and those events need to be interpreted.

As Soller et al. (2005) describe, technological tools can be distinguished according to which agent – the tool, the teacher, or student – is responsible for each step of the process. In the case of teacher noticing in the context of teacher dashboards, the agents are the teacher and the dashboard, and the tasks over which responsibility can be divided are the detection of relevant events and the interpretation of those events. The first scenario is that of *mirroring* dashboards. In this case, data about learners is collected in the digital learning environment and analyzed by the system. The dashboard shows the resulting information to the teacher and he or she can peruse it at his or her own discretion. Paying attention to relevant information and interpreting it is left to the teacher. Examples of mirroring dashboards in the context of CSCL are the work by Melero et al. (2015), Schwarz and Asterhan (2011), and Van Leeuwen et al. (2014, 2015b). Van Leeuwen et al. (2014) for example visualized the amount of effort put in by each group member. The teacher could consult this information at will, without the system prompting or alerting the teacher to do so. The second level of aid is that of *alerting* dashboards. Besides displaying information, the teacher dashboard also provides alerts or classification of groups that need attention by comparing the groups' status to some standard. Kaendler et al. (2016) showed that teachers are indeed in need of support of knowing what information about collaborating students to look for. Examples of alerting teacher dashboards in the context of CSCL include the work by Casamayor et al. (2009), Gerard and Linn (2016), Martinez-Maldonado et al. (2015), Segal et al. (2017), Schwarz et al. (2018), and Voyiatzaki and Avouris (2014). Schwarz et al. (2018) for example developed a system that informs teachers of predefined critical moments so that the teacher can decide whether and how to act. Finally, *advising* dashboards not only display information and alert to certain events, but also assist in the interpretation of the information of which the teacher is alerted by providing additional advice about what is happening or what action the teacher could undertake. We could only find one example of an advising dashboard in the context of CSCL, namely the work by Chen (2006), in which the system generates examples of support that a teacher could offer to students once an important event had been detected.

Descriptive studies show that teachers find mirroring dashboards useful to gain insight into and understand the development of student understanding of task material during CSCL (Melero et al. 2015; Schwarz and Asterhan 2011). However, in experimental studies, teachers do not always show improved detection and interpretation of relevant events of collaborating students with mirroring teacher dashboards (Van Leeuwen et al. 2014, 2015b). As these studies show, providing teachers with information (i.e., mirroring) can lead to different types of inferences, and not always to a correct prioritization of which group is most in need of help. A common thread seems to be that teachers do experience more confidence in their judgement of whether student support is necessary when they are provided with a dashboard, as it acts as an additional source of information about their students. Martinez-Maldonado et al. (2015)

compared a mirroring to an alerting teacher dashboard and found that only in the alerting condition, the teachers' feedback significantly influenced students' achievement. It could mean that teachers were better able to detect relevant information, upon which they could successfully act. In line with this finding, Casamayor et al. (2009) found that teachers detected more events when they were provided with a CSCL teacher dashboard. The study by Chen (2006) does not provide much detail about teachers' evaluation of the advising dashboard. It is therefore yet unclear whether advising dashboards enhance teachers' interpretations of collaborative situations.

It must be noted that the amount of experimental research that investigates the effect of teacher dashboards with these different levels of aid is scarce. In a review of teacher tools, Sergis and Sampson (2017) confirm that only little attention has been paid to especially the higher levels of interpretational aid. They hypothesize that merely providing information to teachers (i.e., mirroring) might not be enough, and that teachers need advice or recommendations (i.e., advising) for how to translate data to specific insights about student activities. To the best of our knowledge, there has not yet been a systematic comparison of mirroring, alerting, and advising teacher dashboards within a single experimental study. The goal of the present paper is therefore to investigate the process of teacher noticing in the context of CSCL teacher dashboards. More specifically, we zoom in on how teachers interpret the dashboard and on teachers' cognitive representation of dyads once they have seen the dashboard, and how this is influenced by the function the dashboard fulfills. Our goal is to test whether more aid indeed helps to detect and interpret information on a teacher dashboard.

The present study

As it is essential teachers are informed of students' activities, we examine how teachers interpret teacher dashboards that provide information about cognitive and social aspects of the activities of collaborating students in the domain of fractions. Three versions of a dashboard were created to experimentally investigate three functions of the dashboard: a dashboard that provides information (mirroring), a dashboard that provides information and aids detection of relevant information (alerting), and a dashboard that provides information, aids detection, and aids interpretation of relevant information (advising). In the context of CSCL, where teachers need to make a quick succession of decisions, it is essential that teachers find relevant information quickly and accurately and can make the right inferences about their students in order to stimulate effective collaboration. In the present study, we therefore focus both on the time teachers need to make sense of teacher dashboards, and also on the underlying processes of detection and interpretation of information, which have only rarely been studied in detail in the field of CSCL. The study is performed in the context of student collaboration on fraction assignments. The following research questions were formulated for the experimental study:

What is the influence of mirroring, alerting, and advising CSCL teacher dashboards that display information about student collaboration on:

- *the speed with which teachers detect and interpret information?*
- *teachers' detection of relevant information?*
- *teachers' depth of interpretation of information?*
- *teachers' experienced cognitive load and confidence in detecting and interpreting information?*

In the alerting and advising conditions, aid is provided for detecting or for detecting and interpreting information about CSCL situations, respectively. We therefore expect that: *Hypothesis 1*: teachers in the alerting and advising condition need less time for the process of detecting and interpreting the dashboards than teachers in the mirroring condition.

Because aid for detecting relevant information is provided in both the alerting and advising condition, we furthermore expect that *Hypothesis 2*: teachers in the alerting and advising condition more accurately detect in which group and what type of event occurred than teachers in the mirroring condition, that they are more confident of their selection and need less effort to select a group.

Lastly, because aid for interpreting information is given in the advising condition, we expect that *Hypothesis 3*: Teachers in the advising condition are capable of a richer interpretation of the teacher dashboards than teachers in the other two conditions, and are more confident of their interpretation and need less effort to interpret the dashboard.

In the first phase of the research, a teacher dashboard prototype was created. Teacher interviews were held to determine the type of information that teachers would find informative to aid teacher monitoring of collaborating students. Subsequently, three versions of this prototype were created to compare the mirroring, alerting, and advising function. In the next section, we first describe more details about the context of the study and the initial teacher interviews that led to the development of the dashboard prototype. We then continue with a description of the experimental study.

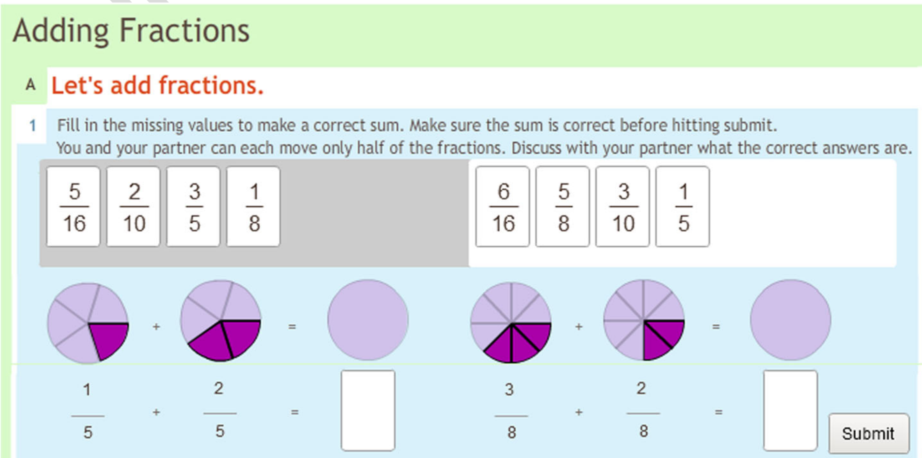
Development of the CSCL teacher dashboard (prototype) used in this study: Pilot interviews with teachers

The teacher dashboard prototype in the present paper was developed in the context of collaborative learning of fraction assignments, with students from 3rd and 4th grade (primary education) collaborating in dyads. Fractions are an essential basic skill that students need to develop to prevent misconceptions that can hinder mathematical ability at later ages (Booth and Newton 2012; Bailey et al. 2012). Siegler et al. (2013) describe that in particular students' conceptual knowledge about fractions and their attentive behavior to the task predict gains in fraction skills. CSCL can be a valuable tool in stimulating both practice of fractions and attention to the task. Through collaboration, the interaction between learners is assumed to contribute to understanding of the material. Simultaneously, by using a computer-supported environment, part of the collaboration can be guided and attention to the task be stimulated (Stahl et al. 2006). The program MathTutor (2018) is specifically designed as a support system for CSCL in the math domain (including fractions), and thereby combines the advantages of a tutoring system and CSCL (Olsen et al. 2014). It also specifically targets collaboration between primary school students, which is less common than studies investigating collaboration between older students. MathTutor allows collaborative practice of both conceptual and procedural knowledge of fractions, and has been shown to be equally effective to individual practice in terms of learning gains, with the collaborative setting showing faster progress (Olsen et al. 2014). For the purpose of this specific paper, we report about the skills of naming fractions, simplifying fractions, and adding and subtracting fractions.

When dyads collaborate on fractions tasks in MathTutor, each member of a dyad has his or her own computer screen, but the interface is the same for the two students. Students are seated next to each other and can discuss the assignments by talking aloud. Each action of one of the group members is visible to the other, while each member can manipulate particular parts of the interface. An example of an assignment students may work on in MathTutor (see Fig. 1) is to complete the sum of two fractions by inputting the correct outcome (in the two white squares in the lower half of the screen), which can be chosen from several answer options (the grey and white area in the upper half of the screen). In this case, the two members of the dyad can each drag and drop the fractions from either the grey or the white area to complete the sum in the lower part of the screen. Thus, student 1 is able to drag answer options that are in the grey area to the lower part of the screen, while student 2 has control over the answer options in the white area. As answers may be necessary from both the grey and the white area, input from both members is required and attention to the task from both members is stimulated.

It can take dyads multiple attempts to correctly solve an assignment. Dyads can check their answers for correctness and MathTutor also enables step-based guidance by providing specific hints. Once the current assignment is correctly solved, dyads move on to the next assignment. Because each assignment is coupled to a specific fraction skill, MathTutor also tracks the development of students' proficiency in each skill. Thus, when MathTutor is employed in the classroom, the technological platform provides students with a basic support structure, while essential higher-order support is offered by the teacher (Saye and Brush 2002; Slotta et al. 2013). Because MathTutor tracks all student activity, these digital traces could be used as input for the development of a teacher dashboard.

To ensure high usability of the dashboard, a teacher co-design methodology was employed (Matuk et al. 2016). Elaborate interactive sessions with 10 primary school teachers were held. They had a mixed background concerning experience with CSCL learning environments (Van Leeuwen and Rummel 2018). Here, we report the findings that were relevant to the development of the teacher dashboard. We first asked teachers what types of events they monitor or act upon in the situation of monitoring a class of collaborating dyads. We specifically made use of contextual inquiry (Hanington and Martin 2012) so that teachers did not indicate what information they would like to have, but what they actually do in the classroom. Teachers indicated five types of information as sources of information or types of events that they act upon in the collaborative classroom: 1) background information about students, such as prior performance in math and whether the dyad was a good fit in terms of good collaborators, 2) information about dyads' grasp



Q1

Fig. 1 Screenshot of a fraction assignment of the materials and whether dyads get stuck on a task, for example when the task is too difficult, 3) dyads' involvement or engagement with the task, 4) information about dyads' progress on the task and whether students understand why they made progress (or not), and 5) information about potential difficulties with interaction between group members.

Aside from background information about students, all types of information could be mapped onto the distinction that is often made in literature between cognitive, metacognitive, and social aspects of collaboration (Kaendler et al. 2015). We therefore aimed to include information about these three aspects on the teacher dashboard prototype. Our plan for the prototype was to test it with simulated, fictitious CSCL situations in which teachers would not know the students beforehand. We therefore did not use background information about dyads in the subsequent investigation.

Our next step was to present teachers with specific possible sources of information in the context of MathTutor, that is, to present the indicators that MathTutor automatically logs about student activity. We thereby aimed to uncover which indicators of student collaboration would be most useful to teachers to decide whether groups faced one of the issues identified by the teachers above. We used the Kano method (Hanington and Martin 2012) to determine which indicators had highest priority/usability. The Kano method stems from the field of product development and is used to discover which features of a product or service are most likely to lead to customer satisfaction (Witell et al. 2013). In this case, teachers were presented with a list of possible indicators of student collaboration, for example "the number of tries a dyad needs to solve an assignment", and were asked to reflect on what the availability of this indicator would mean for the quality of their guidance of the classroom (in terms of having a positive, neutral, or negative effect), as well as what the *unavailability* would mean. We preselected a list of possible indicators to avoid teachers being unaware of some of the possibilities. If the availability of an indicator was judged to have a positive influence and the unavailability a negative influence, the indicator is assumed to be important. On the other hand, when the availability has a positive influence and the unavailability has a neutral influence, the indicator is assumed to be desirable, but not essential. The following four indicators were judged most highly: 1) the number of tries a dyad needs to solve an assignment, 2) the chance that a dyad displays trial-and-error behavior on an assignment, 3) dyads' proficiency on fraction skills, and 4) a display of a dyads' activity over time. These four indicators were selected as input for the dashboard prototype.

One particular indicator that MathTutor does not yet track, is the amount of talking a dyad engages in while solving an assignment. Because we assumed the amount of talk would be indicative of students' engagement in the task as well as a condition for collaborative discourse to occur, we decided to add this indicator to the dashboard as well. Although in its current form MathTutor is not yet capable of doing so, it is not hard to imagine that in future versions the computer's microphone would at least allow for tracking how much sound or talk is generated students generate during CSCL (e.g., Grawemeyer et al. 2017). Informing teachers of the *content* of students' talk would most likely have been even more beneficial, but automatic detection of content of speech is much more complicated to achieve (Shadiev et al. 2018).

The interviews thus yielded the information about student collaboration in MathTutor that was used as input for the teacher dashboard prototype. To inform future research on implementation of such a dashboard in the classroom, the objective of the present study was to investigate the process of teacher noticing of collaborative situations and how this process is influenced by the amount of aid the dashboard provides in detecting and interpreting information. We therefore created three versions of the dashboard prototype: a dashboard that provides information (mirroring condition), a dashboard that also aids detection of relevant information (alerting condition), and a dashboard that aids detection and interpretation of relevant information (advising condition). In an experimental environment, the dashboards were shown to teachers making use of simulated CSCL situations. By doing so, we could investigate the role of the dashboard function on teacher detection and interpretation of information on the teacher dashboards.

Method of the experimental study

Design and participants

An experimental study with a between-subjects design with three conditions was conducted in which participants were asked to detect and interpret relevant information about CSCL situations in a fictitious class displayed on one of three types of prototype teacher dashboards. The three dashboard types differed in the type of support they provided to teachers: 1) mirroring, 2) alerting, or 3) advising. We investigated whether dashboard type influenced speed, accuracy, and depth of detection and interpretation of information on the dashboard.

The sample consisted of 53 participants, 4 of whom were male. The sample did not include any of the teachers that took part in the co-design phase of the dashboard described above. Participants signed up for the experiment voluntarily and received a monetary compensation for their participation. All participants were either pre-service primary school teachers or primary school teachers who had recently finished their teacher education. The majority of female teachers in the study's sample is representative of primary school teacher populations in many countries (World data bank 2018), including the Netherlands (83.9% in 2016; see STAMOS 2018), where the present study was conducted. In Dutch primary school classrooms, collaborative learning is increasingly encountered as part of the standard curriculum (European Parliament 2015), and collaborative software is increasingly implemented (Kennisnet 2015).

As participants differed in the amount of prior teaching experience, we first sorted them into six experience groups (0–10 months, 10–20 months, 20–30 months, 30–40 months, 40–50 months, or 50–60 months teaching experience). Within those groups, participants were randomly distributed over the three conditions in the experiment. Table 1 shows participants' demographics for each condition. No significant differences between the conditions were found regarding age ($F(2,50) = .99$, $p = .38$) or teaching experience ($F(2,50) = .08$, $p = .93$). None of the participants had experience with the MathTutor software that the experiment was based on.

Materials - questionnaires

To make sure the three experimental groups did not differ concerning background characteristics, a number of questionnaires were administered before completing the dashboard trials. Both teachers'

Table 1 Demographics for each experimental condition

	Mirroring	Alerting	Advising
n	18	17	18
Number of males	2	2	0
Mean age in years (SD)	20.2 (2.2)	21.1 (2.7)	21.1 (1.6)
Mean months of teaching experience (SD)	22.0 (18.6)	24.3 (19.4)	24.0 (19.0)

pedagogical experiences and beliefs were taken into account given that these areas are known to possibly play a role in teaches' interaction with technology (Admiraal et al. 2017).

Participants indicated to what extent they had experience with implementing collaborative learning in their classroom and with teaching fractions. Both were multiple choice questions with the options 0 lessons, 1–5 lessons, 6–10 lessons, 11–15 lessons, and more than 15 lessons. They also indicated whether they had experience with teacher dashboards during teaching (yes/no), and if so, with which program.

We also measured teachers' beliefs about the importance of specific student activities during collaborative learning. Teachers were asked to judge how much importance they ascribe to several aspects of student collaborative behavior. Derived from Kaendler et al. (2016), we created items for cognitive and social aspects. Furthermore, we measured teachers' beliefs about their own role during students' collaborative activities, making a distinction between a primarily teacher-regulated process (external regulation) and a primarily student self-regulated process (internal regulation), adjusting the scales by Meirink et al. (2009) to fit the collaborative context. Analyses showed that all four scales had low reliability (Cronbach's alpha lower than .70 in all cases), so we decided to not make use of these questionnaires in subsequent analyses.

Feelings of self-efficacy in the domain of teaching with technology (Tech-SE) were measured using the 7-item scale from Admiraal et al. (2017). An example item was "I have sufficient knowledge to apply ICT in my teaching activities". Cronbach's alpha was .89.

Materials – Dashboard trials

Prior to the dashboard trials, participants received an introduction to the study and an instruction about the specific task. Participants were asked to imagine they were in a classroom situation in which they were a substitute teacher in a 3rd or 4th grade class, where they had just given instruction to the students about fractions, and had instructed students to collaborate in dyads on fraction assignments. Because no immediate questions arose from the students, the teacher decided to consult the teacher dashboard to see if any group was encountering a problem and might need help. The participants' task was thus to observe the dashboard as if they were in this classroom situation, and to find out whether any of the collaborating groups was facing a problem. Participants were told that the main task in the experiment consisted of interpreting 8 dashboard situations that visualized data from fictitious classrooms consisting of 5 dyads. Although the dashboard situations were fictitious, it was stressed in the plenary explanation that participants should try to imagine they were in an actual classroom and therefore try to complete each trial as quickly and accurately as possible, so that as in a real situation, they could close the dashboard and turn back to their classroom again as quickly as possible. By presenting each participant with the same 8 fictitious classroom situations, the effect of different types of teacher dashboards could be systematically investigated. A similar methodological approach was used by for example Chounta and Avouris (2016), Mazza and Dimitrova (2007), and Van Leeuwen et al. (2014, 2015b).

Every dashboard situation started with a short description of the specific context, including whether it concerned a 3rd or 4th grade class, and whether the group had already spent time on fractions. Then, the actual dashboard was shown, on which the participants could find information about the fictitious classroom. The dashboards were static in the sense that the displayed data did not change while the participants looked at the dashboards.

From the initial teacher interviews, five indicators were identified on which information was displayed on the dashboard, namely 1) the number of attempts a dyad needed to solve an assignment, 2) the chance that a dyad displayed trial-and-error behavior on an assignment, 3) dyads' proficiency on fraction skills, 4) a display of a dyads' activity over time, and 5) the amount of talk for each dyad member. Because most indicators are directly linked to individual assignments (for example, the number of attempts a dyad needed to solve per assignment), we thought that it was also necessary to display how many assignments each dyad had already solved. The number of solved assignments thereby constituted the sixth and final indicator.

Figure 2 shows the "start screen" of the dashboard. On the left, the dyad numbers are displayed. On the top row, the six indicators are displayed. When a group number is clicked on, a group overview opens that displays information on all six indicators for that particular group (see Fig. 3). When an indicator is clicked on, a class overview opens that displays information concerning that indicator for each of the five groups.

Figure 3 shows an example of a group overview. On the left, a graph with assignment number on the horizontal axis displays whether an assignment was completed (1), how many attempts the dyad needed on the assignment (2), whether there was a chance the dyad displayed trial-and-error behavior (3), and how often each group member talked while working on the assignment (4). The chance of trial-and-error behavior was based on the speed of activity of a dyad in combination with the number of attempts needed to complete an assignment (high speed and high number of attempts indicating a higher chance at trial-and-error behavior). The amount of talk was based on the sound each student's laptop detected, and thus did not offer information about what was being said. The three colors within the graph display the three fraction skills that the assignments belong to. On the right of the group overview, three bars display the dyads' proficiency at the three fraction skills (5), and below that, a timeline shows the dyads' activity from the start of the lesson until now (6). The dyads' proficiency was based on the number of completed assignments in combination with the number of needed attempts for that particular skill. The activity timeline shows a dot whenever one of the group members gives input or clicks a button within MathTutor.

Each dashboard situation consisted of information about the six indicators for the five groups in the fictitious class. To examine whether participants accurately interpreted the information on the dashboards, we created the eight fictitious situations in such a way that they contained a specific problem in one of the collaborating groups. This was done by varying the values on the six indicators in specific ways. To create the problematic scenarios within the dashboard situations, we consulted literature about characteristics of collaborative learning that

Fig. 2 The "start screen" of the dashboard (translated from Dutch)

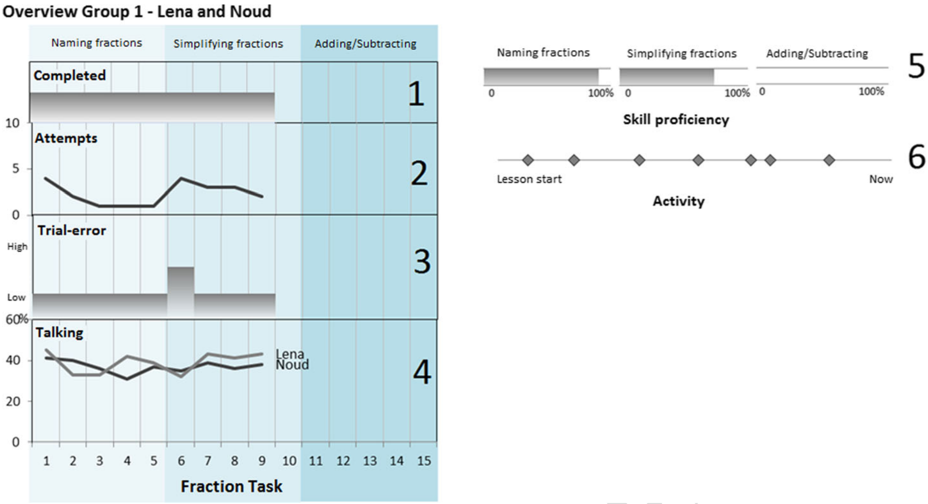


Fig. 3 Example of group overview with 6 indicators of group activity (translated from Dutch), with 1 = completed assignments, 2 = attempts per assignment, 3 = chance of trial-and-error behavior, 4 = amount of talk, 5 = skill proficiency, 6 = activity time line

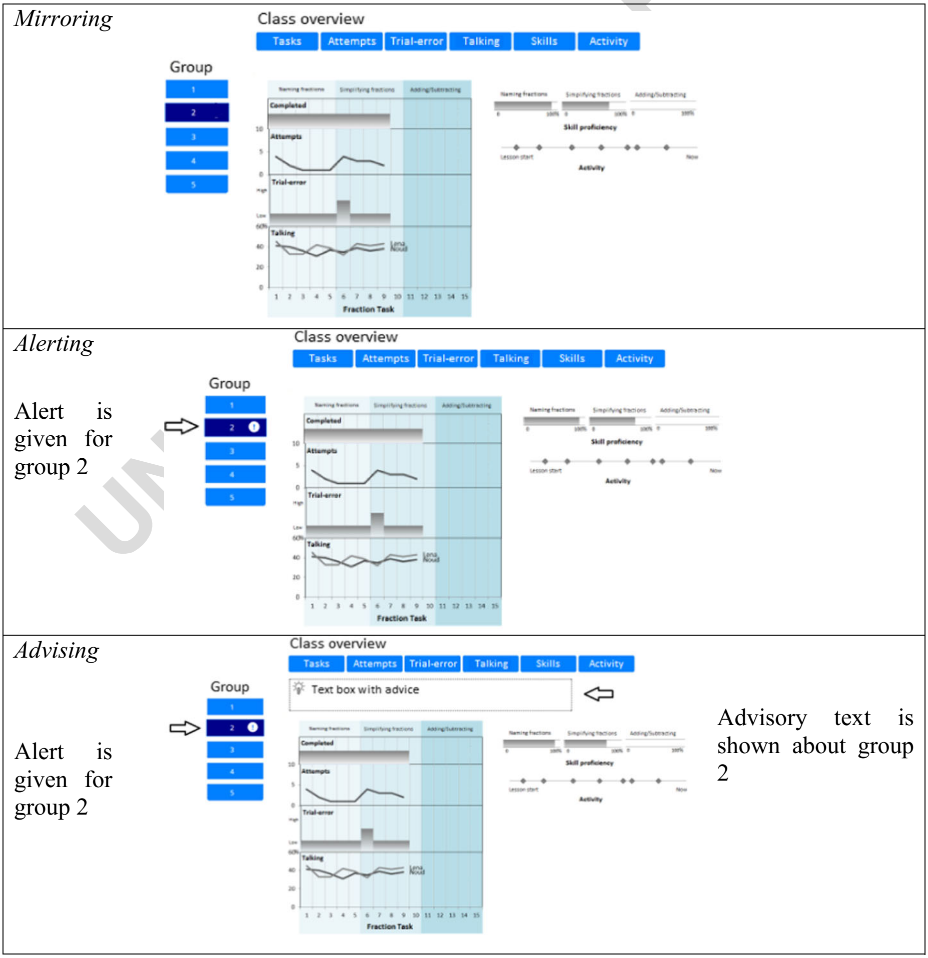
play a role in the success or failure of collaboration (Kaendler et al. 2015; Kahrimanis et al. 2011; Meier et al. 2007), as well as the findings from the teacher interviews. We used the distinction between cognitive and social aspects to categorize the problematic situations. Because it concerned static, short CSCL situations, we decided to not include metacognitive problems such as students lacking insight into the progress they are making. Table 2 displays the six problematic situations that were created. Two problems concerned cognitive aspects of collaboration, two problems concerned social aspects, and two problems were a combination of a cognitive and social problem. For each problem, we created a dashboard situation in which one of the five groups experienced this problem by setting up the values on the six indicators in a particular way. For the remaining four "unproblematic" groups in a dashboard situation, the values on the indicators were kept average. Furthermore, two dashboard

Table 2 Overview of dashboard situations

Situation	Type	Problem within one dyad
1	Cognitive	The dyad is stuck on an assignment concerning adding fractions because they have not adequately mastered the needed skill of equalizing fractions.
2	Cognitive	The dyad shows off-task behavior.
3	Social	One student within the dyad is dominating the collaboration.
4	Social	The two members of the dyad take turns solving the assignments instead of collaborating on the assignments.
5	Cognitive and Social	The dyad is stuck in an argument, and because they disagree, they are stuck on the assignment.
6	Cognitive and Social	The dyad displays trial-and-error behavior, which means they are not discussing the task and are not gaining understanding of the material.
7	No problem	None
8	No problem	None

situations in which no problem occurred were included, so situations in which all five groups performed more or less on an average level.

Participants in the three conditions were presented with the same fictitious classroom situations, but the amount of help that the dashboard provided in detecting and interpreting the problem that occurred in one of the groups, differed. Figure 4 depicts screenshots from the three different types of teacher dashboards that constituted the 3 study conditions: mirroring, alerting, and advising. In the mirroring condition, the dashboard displayed information about each collaborating group that the teacher could view upon demand. In the alerting condition, the dashboard display was enhanced with an exclamation mark on the button of one of the groups, thereby alerting them to groups of collaborating students that deviated in some way from the other groups. In the advising condition, alerts were given as well, and if the teacher opened the corresponding group overview, a light bulb with a supporting prompt advised teachers about how to interpret the situation in the group for which an alert was given. The advice followed the same format in each situation. First, a statement was given about the type of problem the group seemed to be having (e.g., "This group seems to have a cognitive



Q2

Fig. 4 Screenshots of mirroring, alerting, and advising dashboard (translated from Dutch) problem”). Then, a more specific explanation was given of what the problem might be, making reference to the indicators on the dashboard (e.g., “John and Emma need more attempts on the tasks as time progresses. They do not seem to be proficient at simplifying fractions, because of which they get stuck at adding and subtracting.”). In the two situations where no problem occurred, neither alerts nor advice were present.

While interacting with the teacher dashboard, teachers could open group and class overviews as often as they liked. Only one overview could be opened at once. A green button was available in the bottom right corner (see Fig. 2) that could be pressed once the teacher was done interpreting the situation. Once they pressed ‘Finished’, participants answered four questions about their interpretation of the situation.

The first question was which of the five groups had faced a problem. Participants could choose one of the five groups, or select the option ‘none of the groups’ (so they could not select multiple groups). Participants were also asked how much effort it had taken them to determine the answer, and how confident they were of their answer. The amount of effort, which can be regarded as an indicator of experienced cognitive load, was measured using the scale developed by Paas (1992), ranging from 1 (very, very little effort) to 9 (very, very much effort). The confidence question was measured on a scale from 1 (very unsure of my answer) to 10 (very sure of my answer).

If participants selected the option ‘none of the groups’, they were able to explain their answer in an open comment, but were not asked any subsequent questions. If participants indicated there was indeed a problem within one of the groups, they were asked in a second multiple-choice question to categorize the problem as either cognitive, social, or both (in line with the terminology that was used in the advice boxes in the advising condition). This distinction between cognitive and social aspects was based on the framework provided by Kaendler et al. (2015) that we discussed earlier. Metacognition was not included as a problem category as we had not included any situations of this type. In the answer options of the multiple-choice question, each answer option was accompanied by a short explanation of what we meant by each category (e.g., “Cognitive – the problem is related to (understanding of) the task content”). Furthermore, participants were asked to explain in an open comment as fully as possible why they thought it was this type of problem and what they would do as teachers in this situation. For these two questions together – problem type and explanation – participants were again asked to indicate the associated amount of effort and confidence level.

Procedure

Participants could sign up for a timeslot to take part in the study distributed over several weeks, which meant a differing number of participants took part in each session. The tables with computers were separated by screens so that there was no contact between participants during data collection. Table 3 outlines the general procedure of the experiment. After an explanation of the context and procedure of the experiment, participants signed informed consent. They then opened the experimental software (Gorilla software 2018) through a login code that was coupled to their assigned condition. The experiment started with filling in background questionnaires. Participants also received a short instruction that was specific to their experimental condition, consisting of screenshots of the dashboard and written explanatory text.

Then, each participant completed 8 dashboard trials in random order. Each dashboard trial was introduced by a short text that explained the situation and reminded participants of the task. Participants could then examine the teacher dashboard prototype for a maximum of

Table 3 Overview of procedure

	Activity	Duration
1	Plenary explanation about the procedure + sign informed consent	10 min
2	Fill in questionnaires: Demographics, including experience teaching fractions, experience with CL, experience with dashboards Technological self-efficacy Teacher beliefs concerning collaborative learning	10 min
3	Condition-specific instruction about dashboard	2 min
4	Dashboard-situations (8x) Description of context Examine dashboard Answer questions about interpretation of situation	30–40 min
5	Fill in questionnaire: general opinion about experiment	2 min

7 min, and once they pressed 'Finished', questions followed about the participants' interpretation of the situation. The next dashboard trial then automatically followed.

Finally, participants were asked about their general opinion of the experiment, including whether the procedure was clear and whether they understood the visualizations on the teacher dashboards.

Data analysis

Each dashboard-trial yielded the following measures: the time needed to interpret the situation (time until 'Finished' was pressed), the selected problematic group (if any), categorization of problem type, and participants' explanation of their interpretation of the situation. Furthermore, invested mental effort and level of confidence associated with each answer was obtained.

The participants' selected group was compared to the group we had intended as the problematic group when designing the fictitious situation, and the amount of accurately selected groups over all eight situations was calculated (ranging between 1 and 8). We also assessed in how many of the 8 situations participants selected the problem type we had intended.

To analyze participants' open comments concerning their interpretation of each situation, a coding scheme was developed based on the teacher noticing framework (Van Es and Sherin 2008). Van Es and Sherin (2008) used a coding scheme to analyze the quality of teachers' interpretations of classroom videos. For the purposes of the current study, we coded the following three dimensions of teachers' interpretations. (1) We coded which of the six indicators on the dashboards the teachers mentioned as *evidence* for their interpretation of the situation (so, which events they had monitored). We used the number of mentioned indicators as a variable of evidence use. (2) Similar to Van Es and Sherin (2008), we coded which *stance* teachers adopted in describing their interpretation. Each comment was coded as either level 1, indicating merely a description of student behavior; level 2, indicating a description and judgement of student behavior, but without argumentation; level 3, indicating description and judgement accompanied by argumentation; or level 4, indicating description, judgement with argumentation, and an argument for why a certain alternative explanation was not applicable. For each participant, we calculated the average score for stance. (3) We coded *specificity* of the interpretation. An interpretation was regarded to be more specific when teachers mentioned a specific fraction skill, a specific student name, a specific number from one of the indicator graphs, a comparison of multiple groups of students, or a comparison of one group over multiple time points. For each aspect, the score for the interpretation specificity increased by 1, with a maximum of 5. The average specificity of interpretations was then calculated per participant. A subset of 85 comments (~25%) was independently coded by the first author and a research assistant to determine interrater reliability. For all three dimensions, more than 80% agreement was established. Therefore, the rest of the data was coded by a single rater and this rater's scores were used in the analyses.

Results

Comparison between conditions and manipulation check

Participants in the three conditions were compared concerning experience with implementing collaborative learning in the classroom (Mirroring: $M = 2.56$, $SD = 1.29$; Alerting: $M = 2.94$, $SD = 1.25$; Advising: $M = 2.89$, $SD = 1.18$), experience with teaching fractions (Mirroring: $M = 1.50$, $SD = 1.46$; Alerting: $M = 2.18$, $SD = 1.29$; Advising: $M = 1.83$, $SD = 1.20$), experience with teacher dashboards (Mirroring: $M = 0.17$, $SD = 0.38$; Alerting: $M = 0.24$, $SD = 0.44$; Advising: $M = 0.17$, $SD = 0.38$), and concerning the score on the Tech-SE scale (Mirroring: $M = 3.79$, $SD = 0.49$; Alerting: $M = 3.77$, $SD = 0.74$; Advising: $M = 4.11$, $SD = 0.61$). No significant differences between conditions were found on any of these background variables, $p > .05$ in all cases.

We also checked whether the procedure and layout of the dashboard trials were clear to participants. In the measurement at the end of the experiment, on a scale from 1 to 5, an average score of 4.5 ($SD = .75$) was found for clarity of the procedure, 4.6 ($SD = .57$) for clarity of the class overviews, and 4.5 ($SD = .61$) for clarity of the group overviews.

Effect of CSCL teacher dashboard type on speed of interpretation

The average time participants needed to evaluate a situation was 59.1 ($SD = 19.3$) seconds in the mirroring condition, 60.3 ($SD = 17.3$) seconds in the alerting condition, and 68.5 ($SD =$ 18.9) seconds in the advising condition. The time needed until 'Finished' was pressed declined in each condition as the number of trials progressed. Where the average lay between 80 and 100 s for the first vignette, and by the eighth vignette the average time had dropped to the range of 40–60 s in all conditions.

A mixed ANOVA with the trials as within subjects factor (with 8 measurement points) and condition as between subjects factor showed that there was indeed an effect of trial number on participants' response time, $F(1,50) = 93.21$, $p < .001$, $\eta^2 = .68$, but no main effect of condition, $F(2,50) = 1.35$, $p = .27$, and no significant interaction between trial number and condition, $F(2,50) = 0.49$, $p = .62$. This finding indicates that all participants needed less time to interpret a situation as the trials progressed, but that there was no statistically significant difference in response time between conditions. Figure 5 shows the development of response time per condition. Looking at the graph, response time in the mirroring and advising condition seems to stabilize, whereas response time in the advising condition continues to drop. It might be that with more trials, the advising condition would have arrived at even lower response times.

We further examined what information participants looked at during vignettes. On average, participants in the mirroring condition clicked on group overviews or indicators 16.31 times ($SD = 4.33$) per vignette, compared to 16.04 ($SD = 3.81$) in the alerting condition and 15.27

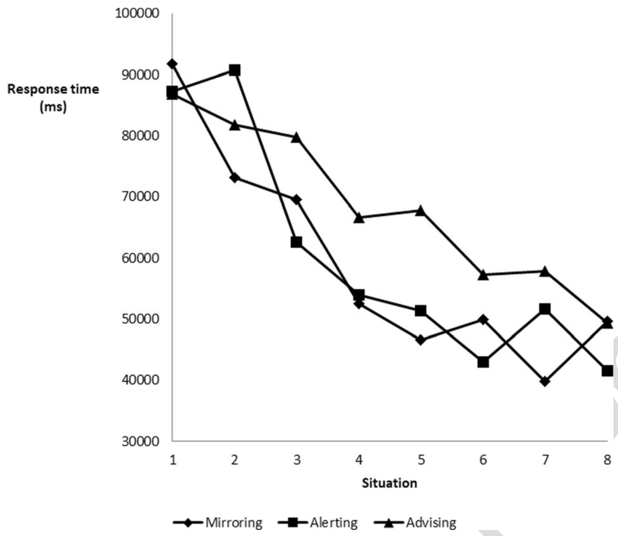


Fig. 5 Development of response time per condition

(*SD* = 3.81) in the advising condition. The number of views did not differ significantly between conditions, $F(2,50) = .32$, $p = .73$.

Out of the five available group overviews, participants in the mirroring condition on average looked at 2.96 group overviews per vignette (*SD* = 1.68), compared to 3.77 groups (SD = 1.44) in the alerting condition and 3.79 groups (*SD* = 1.21) in the advising condition. Again, there was no significant difference between conditions, $F(2,50) = 1.85$, $p = .17$. Interestingly, the mirroring condition looked at the least number of groups, but looked at the highest number of indicators, with an average of 5.70 indicators per vignette (*SD* = 0.45). They thus seemed to have relied more on the indicators than on the group overviews for their judgment of the situations. In the alerting condition, on average 4.24 indicators (*SD* = 1.77) were looked at, and in the advising condition this number was 4.63 indicators (*SD* = 1.67). Analysis showed there was a significant difference between conditions, $F(2,50) = 5.03$, $p = .01$, $\eta^2 = 0.167$. Bonferroni post hoc tests showed a specific difference between the mirroring ($M = 2.96$, $SD = 1.68$) and alerting ($M = 3.77$, $SD = 1.44$) condition, $p = .01$, $d = 0.52$.

Effect of dashboard type on choice of problem group

Over 8 vignettes, the average number of correctly detected problem groups was 6.50 (*SD* = 1.2) in the mirroring condition, 6.82 (*SD* = .95) in the alerting condition, and 7.11 (*SD* = .83) in the advising condition. Although the pattern of the mean scores in each condition was in line with our second hypothesis, with mirroring being lowest and advising being highest, ANOVA showed there was no significant difference between conditions, $F(2,50) = 1.68$, $p = .12$.

The average experienced cognitive load (on a scale from 1 to 9) associated with determining the problem group was relatively low in all conditions, namely 3.65 (*SD* = .87) in the mirroring condition, 3.65 (*SD* = 1.21) in the alerting condition, and 3.07 (*SD* = 1.02) in the advising condition. ANOVA showed there was no significant difference between conditions, $F(2,50) = 1.86$, $p = .17$.

The average confidence level (on a scale from 1 to 10) with which participants selected the problem group was relatively high, namely 7.12 ($SD = .71$) in the mirroring condition, 7.15 ($SD = 1.03$) in the alerting condition, and 7.55 ($SD = .95$) in the advising condition. ANOVA showed there was no significant difference between conditions, $F(2,50) = 1.06$, $p = .35$, although the patterns for both experienced cognitive load and level of confidence were again in line with our expectation (Hypothesis 2).

Effect of CSCL teacher dashboard type on problem type ascribed to chosen group

Next, we examined how often participants selected the problem type (cognitive, social, or both) we had intended in the design of the situations. We first selected those cases in which participants correctly identified the group we had intended to be the problematic group. Of those cases, we calculated what percentage participants selected the problem type we had intended (between 0 and 100%).

Participants on average chose the same problem type as we had intended in 69.5% of cases ($SD = 16.0$) in the mirroring condition, 73.2% ($SD = 15.7$) in the alerting condition, and 76.9% ($SD = 12.1$) in the advising condition. Again, although the pattern was in line with our expectations (Hypothesis 2), ANOVA showed there was no statistically significant difference between conditions, $F(2,50) = 1.17$, $p = .32$.

The average experienced cognitive load associated with determining the type of problem was again generally low, and again showed the expected pattern; 3.64 ($SD = 1.02$) in the mirroring condition, 3.58 ($SD = 1.15$) in the alerting condition, and 3.19 ($SD = 1.04$) in the advising condition. ANOVA showed there was no significant difference between conditions, $F(2,50) = .93$, $p = .40$.

Similar to the findings about selected group, the average confidence level with which participants selected and explained the type of problem was relatively high, namely 7.08 ($SD = .66$) in the mirroring condition, 7.11 ($SD = .98$) in the alerting condition, and 7.21 ($SD = 1.10$) in the advising condition. ANOVA showed there was no significant difference between conditions, $F(2,50) = 0.09$, $p = .91$.

So, although the selected problem *group* (see previous section) in general did match the intended problem group, participants often chose another *type* of problem as we had intended in the design of the dashboard situations. We therefore looked in more detail at the comparison of the intended problem type to the selected problem type, see Table 4. Again, we first selected those cases in which the problem group was accurately detected. This means situations 7 and 8 are not reported in Table 4, as people who had accurately detected the intended group, selected "no problematic group" (and thus no problem type) in these situations. The bold numbers

Table 4 Comparison of intended and selected problem type

Situation	Intended problem type	Selected problem type (%)		
		Cognitive	Social	Cognitive and Social
1	Cognitive	**67.3**	0	32.7
2	Cognitive	**40.8**	20.4	38.8
3	Social	2.3	**88.6**	9.1
4	Social	0	**97.6**	2.4
5	Cognitive and Social	20.8	29.2	**50.0**
6	Cognitive and Social	32.1	5.7	**62.3**

show that in each situation, the predominant selected problem type was the same as the intended problem type. There were also some cases where other problem types were selected. For example, in case of the first two vignettes, which we had intended to display a cognitive problem, in more than 30% of responses, participants indicated there was both a cognitive and social problem.

The results are especially interesting in the advising condition (see Table 5), where the participants were given a text box with a suggestion of the type of problem a particular group was facing. The bold numbers in Table 5 show the predominantly selected problem type in each situation, which was the same to the intended problem type, except in situation 2. Table 5 also shows that a substantial part of participants chose the non-intended problem types in the other situations. These findings could mean that participants in the advising condition to a certain extent disagreed with, adjusted, or ignored the teacher dashboard's suggestion. We discuss some of the participants' comments in these specific cases below.

In situations 1 and 2, most participants agreed there was a cognitive problem, but they often also read a social problem into the situation. Some participants explicitly commented on only partially agreeing with the teacher dashboard. For example (situation 1): "I think the group does not only have trouble in the cognitive domain (as the dashboard suggests), the other assignments went well after all, but that they may also have been chit-chatting or gotten into a conflict."

It was mostly the indicator of amount of talk that the teacher dashboard displayed that led to the distribution in selected problem type. When a dyad displayed more talk than average and the dashboard suggested it had a cognitive cause (situations 1 and 2), participants often adjusted this suggestion to social or affective causes. Encountered explanations included dyads lacking focus or motivation, displaying off-task behavior, and not living up to the responsibility of keeping each other motivated to work on the task. The other way around, when the dashboard suggested a high amount of talk could indicate a social conflict (situation 5), participants suggested it could also mean students were having an engaged conversation about the task material.

As one participant explicated, the amount of talk in itself does not denote the topic of a dyad's conversation: "It [the type of problem] cannot be determined based on this dashboard, as you do not know what the children are saying to each other". Some participants indicated they would like to go observe a dyad to see what was going on specifically. In a real classroom, teachers would of course be able to do so. What is interesting here is the type and amount of possible explanations that the teachers derive from a CSCL situation – the types of "hypotheses" they form about the collaborating students. For example, some participants did weigh several options against each other and counterfeited some problem types. For example, in

Table 5 Comparison of intended and selected problem type in the advising condition

Situation	Intended problem type	Selected problem type (%)		
		Cognitive	Social	Cognitive and Social
1	Cognitive	**77.8**	0.0	22.2
2	Cognitive	16.7	33.3	**50.0**
3	Social	0	**86.7**	13.3
4	Social	0	**100.0**	0.0
5	Cognitive and Social	17.6	5.9	**76.5**
6	Cognitive and Social	16.7	5.6	**77.8**

situation 1, a participant explains the choice for a cognitive problem instead of cognitive and social one: "The description on the dashboard was the same as my own interpretation. The students do not grasp equalizing fractions, seen by the number of attempts and their trial and error behavior. They get stuck because they lack this skill, which they need for adding and subtracting fractions. It is not a social problem, because both students are active and seem to be discussing with each other. I therefore choose a cognitive problem."

To summarize, participants in the advising condition did not just accept the dashboard's interpretation without consideration, and they sometimes adjusted or disagreed with it. Interestingly, a range of interpretations was found that extended beyond the cognitive and social domain, and also included affective aspects of collaboration.

Effect of CSCL teacher dashboard type on depth of interpretation

Participants commented on their interpretation of a situation if they had indeed selected a problematic group. Out of 8 vignettes and 53 participants, 93 times the option "no group had a problem" was selected. That means we obtained a further 331 cases where a group was selected, and where the interpretation was coded. Teachers' interpretations of a situation were coded for four out of eight situations. We selected a situation for each problem type, namely situations 2, 4, 6, and 8, with a total of 160 comments.

Across these four dashboard situations, the average score for each participant was calculated for evidence, stance, and specificity. ANOVA were performed to examine differences between the three conditions, followed by Bonferroni post hoc tests to determine specific differences. ANOVA returned significant differences for evidence, $F(2, 50) = 5.82$, $p = .005$, $\eta^2 =. 189$, and for stance, $F(2, 50) = 5.07$, $p = .01$, $\eta^2 = .17$. Table 6 displays the scores per and comparisons between condition. The mirroring condition showed higher use of evidence than the advising condition, whereas the advising condition showed a higher average stance of interpretation. No significant differences were found concerning specificity of interpretation.

Discussion

As teachers play a large role in CSCL by monitoring and supporting student interaction, it is important they are able to quickly attain an accurate overview of the situation in each collaborating group. Teacher dashboards may help them to do so; however, it is not yet clear which specific role the dashboard should take on in supporting the teacher. In this paper, we investigated teachers' sense making of three different types of teacher dashboards that offered different levels of aid for detection and interpretation of relevant information: mirroring, alerting, or advising dashboards. We thereby took steps in examining and explaining whether teacher dashboards are able to play a role in teacher support of collaborative learning

Table 6 Comparison between conditions for interpretation of situations

Coded aspect	Mirroring	Alerting	Advising	Difference	Cohen's d
Evidence	2.4 (0.7)	2.1 (0.4)	1.8 (0.5)	Mirroring > Advising**	0.98
Stance	2.2 (0.2)	2.5 (0.2)	2.7 (0.3)	Advising > Mirroring**	1.96
Specificity	0.8 (0.2)	0.7 (0.2)	0.6 (0.2)	–	

$^{**}p < .01$

situations, and ultimately, in improving the effectiveness of CSCL. We used a controlled experiment with fictitious collaborative situations in fractions learning to gain more in-depth insight into how teachers go about detecting and interpreting those situations. In the sections below, we discuss our findings and what the findings mean for informing the next steps in researching the implementation of teacher dashboards that enable more effective teacher support of CSCL.

Discussion of findings

Our first hypothesis was that teachers in the alerting and advising condition would need less time for the process of detecting and interpreting information displayed on the teacher dashboards than teachers in the mirroring condition, as they received an alert of which group to look at and, in the advising condition, advice for how to interpret the situation in the group. All conditions showed a decline in time as the situations progressed, which is a sign participants may have needed some time to get used to the layout of the dashboards and to find a routine how to go about interpreting each situation (a common finding in human computer interaction research, Dix et al. 2004). Although we did not find a significant effect of type of dashboard, the advising condition on average needed more time to interpret situations, and their reaction time seemed to stabilize less than in the other two conditions.

With regard to teacher navigation of the dashboards, we found a significant difference concerning the amount of group indicators participants looked at during a situation. Participants in the mirroring condition made more use of the indicators and less of the group overviews, which could be because they did not have help determining which group to visit in the first place. In the other conditions, the participants did not just examine the group they were alerted for, but continued to examine the other groups as well, maybe to check whether they agreed with the dashboard about which group showed a problem. This is why the advising condition on average may have needed longer to interpret situations: they looked at the alerted group and the given advice, in addition to overviews of the other groups. The fact that these participants looked at all available information and not just the singled-out group, means they may have been more likely to also have detected other problems had there been any. Another take on this finding is that participants may not have fully trusted the dashboard's suggestions. The importance of positive attitudes and trust for the adoption of recommendation systems is well researched outside the context of education (Wang and Benbasat 2005), but less so in the context of CSCL teacher dashboards. We further elaborate on this issue below.

We further hypothesized that teachers in the alerting and advising condition would more accurately detect in which group and what type of event occurred than teachers in the mirroring condition, and that teachers in the alerting and advising condition would be more confident of their selection and would need less effort to select a group and problem type. Concerning the detection of the problematic group, we found no significant differences between conditions; all participants seemed to generally select the group we had intended. In contrast to existing studies that did find increased detection ability (Casamayor et al. 2009; Van Leeuwen et al. 2014), our findings complement studies that did not find this effect (Van Leeuwen et al. 2015b). A possible explanation is that the layout of the dashboard in itself, which was created based on a phase of teacher co-design (Matuk et al. 2016), was already a help in detecting the group that faced a problem. The strategy of comparing groups to each other was often employed, which means participants might have been able to single out groups that might need support based on visual cues alone (see Van Leeuwen et al. 2015b, for a similar finding with a mirroring dashboard). So, the phase of detection seemed to be equally effective in all conditions. The generally low levels of experienced cognitive load, high levels of confidence, and high scores on the questionnaire about the clarity of the experiment, indicate that all three versions of the dashboard scored quite high in terms of usability. Teachers were able to detect events with relatively little associated workload, which is in itself a positive indication about the direction we are taking with respect to the development of the dashboard in the specific context of fraction assignments. Of course, it must be noted that the teachers' experienced cognitive load may increase, and the need for dashboard support may therefore become more necessary, in other circumstances that are more similar to the actual classroom. We will return to this issue in the last section, in which we discuss directions for future research.

Concerning the selected problem type, we found no significant differences between conditions either. Interestingly, the choice of problem type deviated more often from what we had intended than the chosen group did. Participants did agree that particular groups faced a problem, but sometimes had different interpretations of what the specific problem was. Teachers seemed to have a specific tendency towards interpreting events as social or affective in nature, so a range of interpretations was found that extended beyond the cognitive and social aspects that we offered as answer options. Potentially, this tendency relates to teachers' beliefs about the importance of social and affective aspects of collaborative learning or their general pedagogical knowledge of what constitutes effective collaborative learning (Gillies et al. 2008; Kaendler et al. 2015; Song and Looi 2012), which unfortunately we were unable to control for due to the low reliability of the employed teacher beliefs scale. Another possible explanation could lie in a lack of relevant domain knowledge concerning fractions, which may have caused difficulty with assessing cognitive problems in the collaborative situations (Park and Oliver 2008; Speer and Wagner 2009). As the advising condition received a text box about the possible problem a group was facing, it is interesting that this condition did not just accept the teacher dashboard's interpretation without consideration, and that they sometimes adjusted or disagreed with the dashboard's suggestion. This finding is in line with the result discussed above, that participants in the advising condition looked at multiple groups' overviews, and not just the overview of the group that they were alerted of. As mentioned, these findings might be related to the amount of trust (Wang and Benbasat 2005) the participants put in the dashboard's suggestions. A study by De Vries et al. (2003) had participants plan a route, either by themselves or by using advice from an automated system. While trust was related to the selection of the automated system, participants also showed a fundamental tendency to trust their own abilities over those of the automated system. An alternative take on the finding that participants looked at a larger number of groups than strictly necessary might lie in teachers' professional identity. Research shows that teachers' professional identity also includes their mode of working with technology (Goos 2005), which can for example mean that teachers view the technology as a 'partner' or as a 'servant'. In the present study, stemming from their particular professional identity, participants may have felt compelled to demonstrate their ability as a teacher by adding value to what a dashboard can do. It is therefore interesting to investigate in future research whether teachers' trust in technology and teachers' professional identity play a role when teachers interact with a teacher dashboard.

The final hypothesis was that teachers in the advising condition would display richer interpretations of the situations than teachers in the other two conditions, and that they would be more confident of their interpretation and need less effort to interpret the dashboards. We can partly confirm this hypothesis, as the advising condition showed a significantly higher stance (Van Es and Sherin 2008) when interpreting the situations than the mirroring condition. We also found that the mirroring condition provided more evidence for their interpretation than the advising condition. Together, it could mean that the mirroring condition focused more on visual aspects of the dashboards and primarily mentioned what they saw (describing stance), whereas the advising condition focused more on what the observation meant (interpreting stance). This finding seems to be in line with the conclusion in the review by Sergis and Sampson (2017) that teachers may have difficulty with interpreting data, and suggests teachers do benefit from support to translate data about collaborating students into an interpretation of the situation. It also fits with our finding that the mirroring condition relied more on the indicators than on group overviews (mentioning indicators constituted the evidence score) when interpreting the situations than the other conditions seemed to do. We found no differences in the scores for confidence or experienced cognitive load associated with an interpretation of a situation. Thus, teachers were equally confident of their judgement, which might mean that teachers are unaware of the depth of their interpretation in terms of the stance they adopt and the amount of evidence they present. Indeed, a large body of work concerns itself with training programs to support teacher development of adequate noticing skills (e.g., Kaendler et al. 2016; Sherin & Van Es, 2008).

To conclude, in terms of teacher noticing of CSCL events making use of teacher dashboards, we can only (partly) confirm our third hypothesis. The level of aid the dashboards offered in detecting and interpreting situations seems to have indeed influenced interpretation of events, but not the detection of events. We therefore provided initial evidence that advising teacher dashboards might be preferable over mirroring or alerting ones, in the sense that teachers gain a higher level of understanding of the situation that may make their subsequent pedagogical actions more effective for supporting CSCL. Further investigation is needed, as the question remains what factors account for this difference in interpretation. We found preliminary indications that teachers in the advising condition spent more time looking at the dashboards, which may mean they took the time to read the dashboard's advice and subsequently contemplated whether they agreed to come to their own conclusion. Supplementary data could shed more light on this question. For example, the earlier mentioned role of teacher beliefs about their own as well as students' role during CSCL could be examined further. Also, measures such as eye tracking or thinking aloud could provide more process data of what teachers look at and how they interpret the data about the situations (Van Leeuwen et al. 2017).

Limitations and directions for future research

The present study and its results must be regarded in light of several limitations and contextual factors. First, given the study's relatively small sample size, and the fact that the study is one of the first to systematically compare several dashboard functions, further studies are required to confirm the results presented here. Second, the study was conducted in a specific context, the characteristics of which could have influenced the results. For example, the study's sample consisted of relatively young teachers. In face-to-face collaboration between students, teachers' amount of teaching experience has been shown to influence the number of times teachers choose to intervene in the groups' work (Goold et al. 2010). It could be that teaching experience also plays a role in *computer-supported* collaborative settings, but this is unknown, as a complicating factor in addressing this question is that experience with technology is likely to interact with these effects as well (Solimeno et al. 2008). Also, the collaborative situations presented here all concerned small groups (of two students) who worked on closed types of tasks in the specific domain of mathematics. Both group size and type of task could be of influence on how students interact, for example by increasing the need for coordinating activities between group members (Chavez and Romero 2012), and subsequently, on how teachers interpret these situations. A specific methodological limitation concerning the domain of the study is that we did not measure participants' mathematical domain knowledge, which has been shown to be important for teaching quality (Hill et al. 2005). Given the complexity of developing and administering such an instrument (e.g., Hill et al. 2008), we used experience with teaching fractions as a proxy. We did try to take into account teachers' pedagogical beliefs, but the instruments we used were not reliable enough to use the acquired data. In short, the results of the present study must be interpreted in light of its specific context, and caution should be exerted when generalizing to other contexts.

Furthermore, the present study was conducted in a controlled setting making use of standardized situations, which differs from the more diffuse and complex situation in the classroom. For example, one could argue that in the classroom, teachers are under time pressure and make use of other knowledge about their students than the information on a teacher dashboard, which means their interpretation of a situation might be more complex as well. On the other hand, the role of these same variables also makes it quite challenging to systematically study the influence of particular types of dashboards on teacher decision making in the classroom. We therefore consider the chosen methodology of a controlled study in which the same collaborative situations could be shown to all teachers, to be a very valuable tool for testing hypotheses concerning teacher noticing processes (and as described, this method is employed by multiple researchers). The methodology allows one to test hypotheses and draw inferences that are controlled for very specific aspects of the situation that might be confounding factors when studying individual, perhaps more authentic, cases. The next step is to examine whether the results can be scaled up to the context of CSCL in an actual classroom. Below, we elaborate on how we could take into account aspects of the real classroom in future studies. We believe each of these factors deserves follow up research, and we discuss how we plan to proceed with next steps.

The first factor concerns time pressure. In the initial situations, teachers took about 80 to 100 s (i.e., ca. 1.5 min) to study the dashboard. It is probably not realistic that the teacher spends such an amount of time on making sense of a dashboard in the classroom, especially when we consider that teachers consult the dashboard multiple times during a collaborative activity (e.g., Schwarz and Asterhan 2011). The question is therefore whether the absence of some expected effects (such as on speed and accuracy of detecting and interpreting information) was caused by the fact that participants could look at the dashboard for such an amount of time. In future studies, time pressure could therefore be manipulated to see whether it affects teacher detection and interpretation of CSCL situations.

A second factor is class size. We employed situations that contained five collaborating dyads, whereas classrooms could contain ten or even fifteen dyads. As the goal of our study was to examine the role of the function of the teacher dashboard, our expectation was that any effect would also be observable with a fewer number of dyads. Future studies could therefore also experiment with increasing the number of collaborating groups and examine whether teachers' speed and accuracy decrease as the number of groups increases (e.g., Chounta and Avouris 2016, who compared 2 versus 4 groups).

This relates to the third factor we did not investigate: the role of teachers' knowledge about their students. To deal with time pressure and large class sizes, teachers partly rely on heuristics to make decisions (Feldon 2007). In the interviews that we performed as a pilot to the experimental study, teachers indicated that they indeed use background information about students to judge a collaborative situation in a classroom, such as prior performance in math and whether the dyad was a good fit in terms of good collaborators. In the fictitious situations, teachers could not make use of this knowledge. Similarly, an additional data source that teachers have in their classroom is to observe their students, interact with them, and thereby check whether their impression of the situation (initially gathered from a dashboard) is correct. As we describe in the Introduction, technological artifacts – including dashboards – have a function within the wider context of the classroom. In this study, we zoomed in on the first step of teachers making sense of the information offered to them through a dashboard. From the teachers' open comments about the situations they were shown, we can gather that teachers were indeed to a certain extent able to pinpoint the type of problem that groups experienced, but also that they would need to observe the students or talk to them to pinpoint the exact problem. Thus, the function of the dashboard could be to support the teacher in obtaining an initial idea of which group to visit and to further initiate teacher-student interaction. Questions for future studies are how teachers respond to situations in which they do know their students, how teachers combine different sources of knowledge (i.e., information from the dashboard and from classroom observations), and what happens when their beliefs about a student conflict with information shown on a dashboard. Again, these are avenues for future research we intend to pursue.

Acknowledgements This work was supported by the Netherlands Organization for Scientific Research (NWO) through a Rubicon grant [grant number 446-16-003/1276]. The authors would like to thank Martijn van der Klis for his technical assistance.

References

Admiraal, W., Louws, M., Lockhorst, D., Paas, T., Buynsters, M., Cviko, A., Janssen, C., de Jonge, M., Nouwens, S., Post, L., van der Ven, F., & Kester, L. (2017). Teachers in school-based technology innovations: A typology of their beliefs on teaching and technology. *Computers & Education, 114*, 57–68.

Booth, J. L., & Newton, K. J. (2012). Fractions: Could they really be the gatekeeper's doorman? *Contemporary Educational Psychology, 37*, 247–253.

Bailey, D. H., Hoard, M. K., Nugent, L., & Geary, D. C. (2012). Competence with fractions predicts gains in mathematics achievement. *Journal of Experimental Child Psychology, 113*, 447–455.

Casamayor, A., Amandi, A., & Campo, M. (2009). Intelligent assistance for teachers in collaborative e-learning environments. *Computers & Education, 53*(4), 1147–1154.

Chavez, J., & Romero, M. (2012). Group awareness, learning, and participation in computer supported collaborative learning (CSCL). *Procedia - Social and Behavioral Sciences, 46*, 3068–3073.

Chen, W. (2006). Supporting teachers' intervention in collaborative knowledge building. *Journal of Network and Computer Applications, 29*, 200–215.

Chounta, I.-A., & Avouris, N. (2016). 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, 21*(4), 815–835.

De Vries, P., Midden, C., & Bouwhuis, D. (2003). The effects of errors on system trust, self-confidence, and the allocation of control in route planning. *International Journal of Human-Computer Studies, 58*(6), 719–735.

Dix, A., Finlay, J., Abowd, G. D., & Beale, R. (2004). *Human-computer interaction* (3rd ed.). Harlow: Pearson.

Duval, E. (2011). Attention please! Learning analytics for visualization and recommendation. In P. Long, G. Siemens, G. Conole, & D. Gašević (Eds.), *Proceedings of the 1st international conference on learning analytics and knowledge* (pp. 9–17). New York, NY: ACM.

European Parliament. (2015). Innovative schools: Teaching & learning in the digital era. Retrieved from http://www.europarl.europa.eu/RegData/etudes/STUD/2015/563389/IPOL_STU(2015)563389_EN.pdf Q4

Feldon, D. F. (2007). Cognitive load and classroom teaching: The double-edged sword of automaticity. *Educational Psychologist, 42*(3), 123–137.

Gerard, L. F., & Linn, M. C. (2016). Using automated scores of student essays to support teacher guidance in classroom inquiry. *Journal of Science Teacher Education, 27*(1), 111–129.

Gillies, R. M., Ashman, A., & Terwel, J. (Eds.). (2008). *The teacher's role in implementing cooperative learning in the classroom*. New York, NY: Springer.

Gillies, R. M., & Boyle, M. (2010). Teachers' reflections on cooperative learning: Issues of implementation. *Teaching and Teacher Education, 26*(4), 933–940.

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.

Goos, M. (2005). A sociocultural analysis of the development of pre-service and beginning teachers' pedagogical identities as users of technology. *Journal of Mathematics Teacher Education, 8*(1), 35–59.

Gorilla software (2018). www.gorilla.sc/about. Accessed at March 7, 2018.

Grawemeyer, B., Mavrikis, M., Mazziotti, C., Hansen, A., Van Leeuwen, A., & Rummel, N. (2017). Exploring students' affective states during learning with external representations. In *Proceedings of the international conference on artificial intelligence in education* (pp. 514–518).

Greiffenhagen, C. (2012). Making rounds: The routine work of the teacher during collaborative learning with computers. *International Journal of Computer-Supported Collaborative Learning, 7*(1), 11–42.

Hanington, B., & Martin, B. (2012). *Universal methods of design: 100 ways to research complex problems, develop innovative ideas, and design effective solutions*. Osceola, US: Rockport Publishers.

Hill, H. C., Loewenberg Ball, D., & Schilling, S. G. (2008). Unpacking pedagogical content knowledge: Conceptualizing and measuring teachers' topic specific knowledge of students. *Journal for Research in Mathematics Education, 39*(4), 372–400.

Hill, H. C., Rowan, B., & Loewenberg Ball, D. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal, 42*(2), 371–406.

Hoogland, I., Schildkamp, K., Van der Kleij, F., Heitink, M., Kippers, W., Veldkamp, B., & Dijkstra, A. M. (2016). Prerequisites for data-based decision making in the classroom: Research evidence and practical illustrations. *Teaching and Teacher Education, 60*, 377–386.

Kaendler, C., Wiedmann, M., Rummel, N., Leuders, T., & Spada, H. (2016). Monitoring student interaction during collaborative learning: Design and evaluation of a training program for pre-service teachers. *Psychology Learning & Teaching, 15*(1), 44–64.

Kaendler, C., Wiedmann, M., Rummel, N., & Spada, H. (2015). Teacher competencies for the implementation of collaborative learning in the classroom: A framework and research review. *Educational Psychology Review, 27*(3), 505–536.

Kahrimanis, G., Chounta, I.-A., & Avouris, N. (2011). Rating dimensions of collaboration quality in synchronous collaborating dyads: Findings and interpretations. In *Proceedings of computer supported collaborative learning (CSCL) conference 2011* (Vol. 1).

Kennisnet (2015). Vier in balans-monitor 2015. Retrieved from http://archief.kennisnet.nl/fileadmin/contentelementen/kennisnet/Onderzoek/Documenten/200108_Vier_in_Balans/Kennisnet_Vier_in_balans-monitor_2015.pdf.

Kyndt, E., Raes, E., Lismont, B., Timmers, F., Dochy, F., & Cascallar, E. (2014). A meta-analysis of the effects of face-to-face cooperative learning. Do recent studies falsify or verify earlier findings? *Educational Research Review, 10*, 133–149.

Lang, C., Siemens, G., Wise, A. F., & Gasevic, D. (2017). *Handbook of Learning Analytics*. Society for Learning Analytics Research.

Martinez-Maldonado, R., Clayphan, A., Yacef, K., & Kay, J. (2015). MTFeedback: Providing notifications to enhance teacher awareness of small group work in the classroom. *IEEE Transactions on Learning, 8*(2), 187–200.

MathTutor (2018). Copyright 2009-2018 Carnegie Mellon University, see https://mathtutor.web.cmu.edu/.

Matuk, C., Gerard, L., Lim-Breitbart, J., & Linn, M. C. (2016). Gathering requirements for teacher tools: Strategies for empowering teachers through co-design. *Journal of Science Teacher Education, 27*(1), 79–110.

Matuk, C., Linn, M. C., & Eylon, B.-S. (2015). Technology to support teachers using evidence from student work to customize technology-enhanced inquiry units. *Instructional Science, 43*(2), 229–257.

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.

Meier, A., Spada, H., & Rummel, N. (2007). A rating scheme for assessing the quality of computer-supported collaboration processes. *International Journal of Computer-Supported Collaborative Learning, 2*(1), 63–86.

Meirink, J. A., Meijer, P. C., Verloop, N., & Bergen, T. C. M. (2009). Understanding teacher learning in secondary education: The relations of teacher activities to changed beliefs about teaching and learning. *Teaching and Teacher Education, 25*, 89–100.

Melero, J., Hernández-Leo, D., Sun, J., Santos, P., & Blat, J. (2015). How was the activity? A visualization support for a case of location-based learning design. *British Journal of Educational Technology, 46*(2), 317–329.

Olsen, J. K., Belenky, D. M., Aleven, V., & Rummel, N. (2014). Using an intelligent tutoring system to support collaborative as well as individual learning. *Intelligent Tutoring Systems*, 134–143.

Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive load approach. *Journal of Education & Psychology, 84*, 429–434.

Park, S., & Oliver, J. S. (2008). Revisiting the conceptualisation of pedagogical content knowledge (PCK): PCK as a conceptual tool to understand teachers as professionals. *Research in Science Education, 38*(3), 261–284.

Putnam, R. T., & Borko, H. (2000). What do new views of knowledge and thinking have to say about research on teacher learning? *Educational Researcher, 29*(1), 4–15.

Rummel, N. (2018). One framework to rule them all? Carrying forward the conversation started by Wise and Schwarz. *International Journal of Computer-Supported Collaborative Learning, 13*(1), 123–129.

Saye, J. W., & Brush, T. (2002). Scaffolding critical reasoning about history and social issues in multimedia-supported learning environments. *Educational Technology Research and Development, 50*(3), 77–96.

Schwarz, B. B., & Asterhan, C. S. (2011). E-moderation of synchronous discussions in educational settings: A nascent practice. *Journal of the Learning Sciences, 20*(3), 395–442.

Schwarz, B. B., Prusak, N., Swidan, O., Livny, A., Gal, K., & Segal, A. (2018). Orchestrating the emergence of conceptual learning: A case study in a geometry class. *International Journal of Computer-Supported Collaborative Learning, 12*(3), 1–23.

Segal, A., Hindi, S., Prusak, N., Swidan, O., Livni, A., Schwarz, B., & Gal, K. (2017). Keeping the teacher in the loop: Technologies for Monitoring Group Learning in real-time. In *Proceedings of artificial intelligence in education* (pp. 64–76).

Sergis, S., & Sampson, D. G. (2017). Teaching and learning analytics to support teacher inquiry: A systematic literature review. In A. Peña-Ayala (Ed.), *Learning analytics: Fundaments, applications, and trends* (pp. 25–63).

Shadiev, R., Wu, T.-T., Sun, A., & Huang, Y.-M. (2018). Applications of speech-to-text recognition and computer-aided translation for facilitating cross-cultural learning through a learning activity: Issues and their solutions. *Educational Technology Research and Development, 66*(1), 191–214.

Siegler, R. S., Fazio, L. K., Bailey, D. H., & Zhou, X. (2013). Fractions: The new frontier for theories of numerical development. *Trends in Cognitive Sciences, 17*(1), 13–19.

Siemens, G., & Gašević, D. (2012). Guest editorial - learning and knowledge analytics. *Educational Technology & Society, 15*(3), 1–2.

Slotta, J. D., Tissenbaum, M., & Lui, M. (2013). Orchestrating of complex inquiry: Three roles for learning analytics in a smart classroom infrastructure. In *Proceedings of the third international conference on learning analytics and knowledge* (pp. 270–274). ACM.

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.

Soller, A., Martínez, A., Jermann, P., & 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.

Song, Y., & Looi, C.-K. (2012). Linking teacher beliefs, practices and student inquiry-based learning in a CSCL environment: A tale of two teachers. *International Journal of Computer-Supported Collaborative Learning, 7*(1), 129–159.

Speer, N. M., & Wagner, J. F. (2009). Knowledge needed by a teacher to provide analytic scaffolding during undergraduate mathematics classroom discussions. *Journal for Research in Mathematics Education, 40*(5), 530–562.

Stahl, G., Koschmann, T., & Suthers, D. (2006). In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences Computer-supported collaborative learning: An historical perspective* (pp. 409–426). Cambridge, UK: Cambridge University Press.

STAMOS (2018). http://www.stamos.nl/index.rfx?verb=showitem&item=9.31.2&view=graph. Accessed at march 7, 2018.

Tissenbaum, M., Matuk, C., Berland, M., Lyons, L., Cocco, F., Linn, M., et al. (2016). Real-time visualization of student activities to support classroom orchestration. In C. K. Looi, J. L. Polman, U. Cress, & P. Reimann (Eds.), *Transforming learning, empowering learners: The international conference of the learning sciences (ICLS)* (Vol. 2, p. 2016). Singapore: International Society of the Learning Sciences.

Van Es, E. A., & Sherin, M. G. (2002). Learning to notice: Scaffolding new teachers' interpretations of classroom interactions. *Journal of Technology and Teacher Education, 10*(4), 571–596.

Van Es, E. A., & Sherin, M. G. (2008). Mathematics teachers' "learning to notice" in the context of a video club. *Teaching and Teacher Education, 24*(2), 244–276.

Van Leeuwen, A. (2015). Learning analytics to support teachers during synchronous CSCL: Balancing between overview and overload. *Journal of Learning Analytics, 2*, 138–162.

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.

Van Leeuwen, A., Janssen, J., Erkens, G., & Brekelmans, M. (2015a). Teacher regulation of multiple computer-supported collaborating groups. *Computers in Human Behavior, 52*, 233–242.

Van Leeuwen, A., Janssen, J., Erkens, G., & Brekelmans, M. (2015b). Teacher regulation of cognitive activities during student collaboration: Effects of learning analytics. *Computers & Education, 90*, 80–94.

Van Leeuwen, A., & Rummel, N. (2018). Orchestration tools for teachers in the context of individual and collaborative learning: What information do teachers need and what do they do with it? In J. Kay & R. Luckin (Eds.), *Rethinking learning in the digital age: Making the learning sciences count, 13th international conference of the learning sciences (ICLS) 2018* (Vol. 2, pp. 1227–1234). London, UK: International Society of the Learning Sciences, Inc.

Van Leeuwen, A., & Rummel, N. (in press). Orchestration tools to support the teacher during student collaboration: a review. *Unterrichtswissenschaft.*

Van Leeuwen, A., Van Wermeskerken, M., Erkens, G., & Rummel, N. (2017). Measuring teacher sense making strategies of learning analytics: A case study. *Learning: Research and Practice, 3*(1), 42–58.

Verbert, K., Duval, E., Klerkx, J., Govaerts, S., & Santos, J. L. (2013). Learning analytics dashboard applications. *American Behavioral Scientist, 57*(10), 1500–1509.

Verbert, K., Govaerts, S., Duval, E., Santos, J. L., Van Assche, F., Parra, G., & Klerkx, J. (2014). Learning dashboards: An overview and future research opportunities. *Personal and Ubiquitous Computing, 18*, 1499–1514.

Voyiatzaki, E., & Avouris, N. (2014). Support for the teacher in technology-enhanced collaborative classroom. *Education and Information Technologies, 19*(1), 129–154.

Wang, W., & Benbasat, I. (2005). Trust in and adoption of online recommendation agents. *Journal of the Association for Information Systems, 6*(3), 72–101.

Wise, A. F., & Vytasek, J. (2017). Learning analytics implementation design. In C. Lang, G. Siemens, A. F. Wise, & D. Gasevic (Eds.), *Handbook of learning analytics* (pp. 151–160). Society for Learning Analytics Research.

Witell, L., Lofgren, M., & Dahlgaard, J. J. (2013). Theory of attractive quality and the Kano methodology – The past, the present, and the future. *Total Quality Management and Business Excellence, 24*, 1241–1252.

World data bank (2018). https://data.worldbank.org/indicator/SE.PRM.TCHR.FE.ZS. Accessed at March 7, 2018.

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