



# Studying teacher withitness in the wild: comparing a mirroring and an alerting & guiding dashboard for collaborative learning

Reet Kasepalu<sup>1,2</sup> · Pankaj Chejara<sup>2</sup> · Luis P. Prieto<sup>1</sup> · Tobias Ley<sup>1,3</sup>

Received: 31 May 2022 / Accepted: 6 February 2023 / Published online: 28 December 2023  
© The Author(s) 2023

## Abstract

Teachers in a collaborative learning (CL) environment have the demanding task of monitoring several groups of students at the same time and intervening when needed. This withitness (both the situational awareness and interventions taken in class) of the teacher might be increased with the help of a guiding dashboard alerting the teacher of problems and providing suggestions for interventions. This paper introduces a quasi-experiment carried out in authentic classrooms. We examined how a mirroring and an alerting & guiding dashboard affected the withitness of teachers in a face-to-face learning environment while students discussed and used a collaborative writing tool. Twenty-four teachers were observed, interviewed, and answered surveys in three different conditions altogether: with no extra information about the situation, using a dashboard mirroring low-level data about the collaboration, and additionally an AI assistant indicating problems in pedagogical terms and potential solutions (i.e., a guiding dashboard). The results show that the situational awareness of the teachers increased with the introduction of a mirroring dashboard. The workload of the participating teachers dropped more with the introduction of an alerting & guiding dashboard, helping teachers feel less frustrated and more accomplished.

**Keywords** Artificial intelligence · Teacher withitness · Collaborative learning · Learning analytics · Awareness · Intelligence amplification · Dashboard

## Introduction

Students need to be supported in the learning process to ensure the effectiveness of Collaborative Learning (CL) (Soller et al., 2005). A situationally more aware teacher can commence an intervention to aid the students in their progress, and thus such a teacher

---

✉ Reet Kasepalu  
reetkase@tlu.ee

<sup>1</sup> School of Educational Sciences, Tallinn University, Tallinn, Estonia

<sup>2</sup> School of Digital Technologies, Tallinn University, Tallinn, Estonia

<sup>3</sup> Department for Continuing Education Research and Educational Technologies, University for Continuing Education, Krems an der Donau, Austria

is critical during CL. The teachers' situational awareness in the classroom has been studied in the field of CSCL under the term *orchestration*. Orchestration refers to the way a teacher manages, in real time, a variety of activities using different media and tools (Dillenbourg, 2013). Orchestration of CSCL and its support with dashboards is a key area of investigation regarding the practice of CSCL (van Leeuwen et al., 2019). The concept of orchestration, however, is very broad and complex and hard to operationalise (Prieto et al., 2011). Recent research has proposed the concept of *withitness*, which is the skill of teachers to process and handle classroom situations along two dimensions: classroom awareness (i.e., what is going on in the classroom) and what the teacher is doing to maintain learning (i.e., her interventions, see (Wolff et al., 2020)). Therefore, withitness is the part of orchestration that happens in real-time during the learning activities.

The required effort teachers must invest into the organisation of activities towards orchestration is referred to as orchestration load (Prieto et al., 2018). A perceived heavy orchestration load could have a detrimental effect on the teacher's situational awareness (Kasepalu et al., 2021). Additionally, the lack of knowledge regarding what information to collect could result in the teacher's wasted effort. As an illustration, novice teachers are known to collect too much irrelevant information in class (Wolff et al., 2020), while simultaneously overlooking information that expert teachers identify as significant and explanatory (Gegenfurtner et al., 2020; Wolff et al., 2020). Here a pattern-seeking artificial intelligence (AI) system that provides advice on potential interventions (i.e., a guiding dashboard) could potentially help decrease the load and help teachers to prioritise the investment of their attention (Soller et al., 2005). Researchers have posited that the strengths of the teacher complemented with AI systems may provide even more effective results in enhancing students' learning (Baker, 2016; Holstein et al., 2019b).

AI in the form of a guiding dashboard could aid the teacher in describing the actions happening in the classroom (thus increasing situational awareness) and predict whether the collaboration process is developing into a successful one (i.e., help decide whether intervention is needed). It is known that teachers would like an AI to draw their attention towards groups (or individual students) that need more attention and help from the teacher in CL (Holstein et al., 2019a), to possibly increase classroom awareness and decrease the orchestration load. Here, multimodal learning analytics (MMLA), which captures the multimodal nature of CL, could enable the development of such AI systems. With the use of various data sources (e.g., audio, eye-gaze, video, heart-rate), MMLA captures a more holistic view of collaboration behaviour that allows an understanding of features of the learning activity both from the physical-world along with traces from digital technology. Therefore, an AI system using multimodal data could support the teacher by helping to increase their situational awareness, decrease the workload, and hence support the teacher's decision making in the moment.

However, the state-of-the-art research carried out in (MM)LA on the development of support tools seldom addresses the issue of identifying emergent problems in CL. Furthermore, to the best of our knowledge, there are no LA tools applied to CSCL that help students or teachers flag problems in pedagogical terms or suggest interventions that could be suitable for the particular problem (Worsley et al., 2021). Even more so, studies conducted are often situated in an artificial or training context and not in the authentic classroom. Notwithstanding, providing additional data could overload the teacher in an already complex classroom environment. Orchestration and orchestration load refer to the myriad of constraints and contextual nuances of real classroom practice (Prieto et al., 2011), which makes it difficult to study in controlled settings.

With the purpose of addressing the aforementioned challenges, this study explored how an AI assistant affected the teachers' skill to be "with-it", in other words, to understand the learning situation and proactively react during CSCL. Furthermore, it investigated the way teachers perceived their workload in authentic settings. We used an AI system with three different configurations in the study: firstly, showing no information on CL; secondly, showing multimodal features such as speaking time, turn-taking, writing operation etc. from data collected from CL (mirroring dashboard); and thirdly, alerting the teacher to problems and suggesting an intervention based on the prediction of the current state of collaboration along with the mirroring functionality (alerting & guiding dashboard). We carried out a quasi-experiment to compare the effects of an alerting & guiding dashboard, a mirroring only dashboard and a no information dashboard. This study investigated free-form collaboration (as opposed to, e.g., scripted collaboration, see Kobbe et al. (2007)) in an *authentic* setting, since free-form collaboration is still most widely used in authentic classrooms.

## Related work

### Effective collaboration

We adhere to the understanding that learning is achieved through processes inherent to social interaction: creating shared meaning together (Stahl, 2005; Thalemann & Strube, 2004), in particular, constructing knowledge that none of the collaborators already had prior to the collaboration nor would have been able to construct alone, but instead created together in synergy (Stahl, 2005). Effective CL, however, does not happen spontaneously, and students need to be supported to acquire those skills. Rummel et al. (2011) have designed the Adaptable Collaboration Analysis Rating Scheme (ACARS) with the purpose of being able to measure the quality of collaboration to help find out where additional support for the students is needed (Meier et al., 2007). The rating scheme entails seven dimensions measuring different aspects of collaboration quality: 1) knowledge exchange, 2) sustaining mutual understanding, 3) argumentation, 4) collaboration flow, 5) structure/time, 6) cooperative orientation, and 7) individual task orientation. *Knowledge exchange* draws from the idea that learning during collaboration happens when partners express their ideas and other members externalise theirs (Fischer & Mandl, 2003). But instead of just exchanging information, Thalemann and Strube (2004) add that *sustaining mutual understanding* entails a group building shared knowledge that the students understand in the same way. Furthermore, based on the provided knowledge, students need to make decisions with a critical evaluation of the options, weighing the pros and cons of each option, leading to the third dimension of collaboration quality assessment: *argumentation* (Tindale et al., 2003). Clark's (1996) communication theory teaches us that students need to coordinate the process of their collaboration reported here as *collaboration flow*. The fifth dimension, *structure/time*, requires group members to keep an eye on the allotted time for the tasks and planning their activity in the group (Erkens et al., 2005). The next dimension, which is referred to as *cooperative orientation*, is characterised by a friendly atmosphere in the group, where everyone feels comfortable to share their ideas and the relationships within the group are "symmetrical" (Dillenbourg, 1999). For the final dimension, *individual task orientation*, we investigate the contribution, motivation and interest of each group member, which are assessed separately in accordance with the understanding that a motivated

participant will be focused on the task (Barron, 2000). ACARS can possibly be very useful for collaboration analytics as it gives more insight into causes and potential interventions (see Table 1 for an overview), when compared to a unidimensional assessment of collaboration quality as high or low.

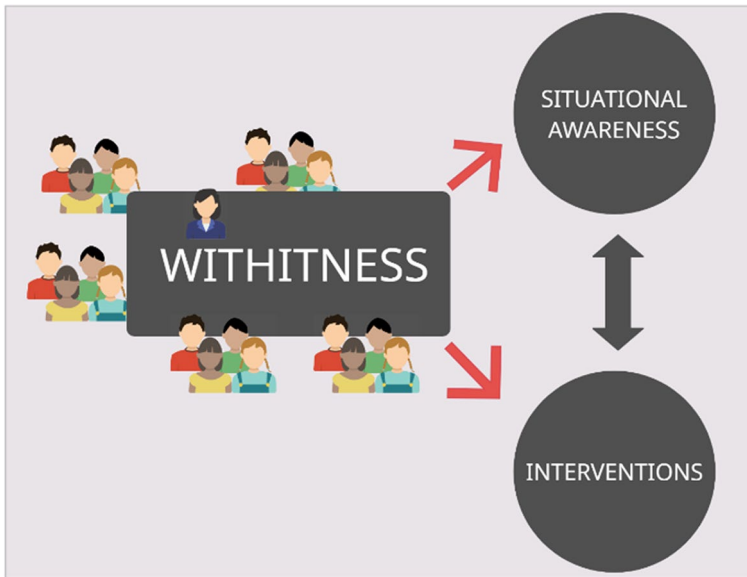
### **Withitness: teacher awareness and intervention in a CL classroom**

In CL, teacher monitoring, decision making during lessons, and interventions carried out have been studied under the term orchestration. When operationalizing orchestration as a term, it is very broad, involving activities prior to the collaboration, during and after (Prieto et al., 2015). In the context of this study, however, we focus on the decisions the teacher makes during learners' collaborative activities. The teacher's skill of noticing, understanding and predicting classroom events and carrying out interventions when needed is called *withitness* (see Fig. 1), which was first coined by Kounin (1970), who made it clear that the teacher needs to constrain poor behaviour in the classroom, and should make a respectable suggestion for amending the behaviour. He also suggested always keeping continual eye contact with students as a means of showing that the teacher is "with-it". Teacher withitness has been studied using video self-analysis (Kounin, 1970; Snoeyink, 2010) and vignette studies (Wolff et al., 2016), either with pre-service teachers (Snoeyink, 2010) or novice and expert teachers (Wolff et al., 2016). Mcdaniel et al. (2009) propose enhancing withitness skills for teachers to be better equipped for teaching in a technology-enhanced learning environment. For the purposes of this study, when we are discussing "withitness", we are focusing less on classroom management issues with the purpose of disciplining the students but are rather concentrating on supporting students to learn to work together as a group, and in particular, to coregulate with the purpose of actively contributing to and benefiting from CL.

As situational awareness is a part of the teacher being "with it", studies show that the classroom awareness of the teachers is increased with the introduction of a LA dashboard in controlled settings (van Leeuwen et al., 2019; Verbert et al., 2014). Endsley (1995) defines situational awareness as the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future. In the classroom context, situational awareness is the teachers' knowledge of what is happening in real-time in the classroom. For the teachers to be "with-it" and make decisions in the classroom, they need to understand and manage classroom situations by first being aware of what is going on in the classroom (situational awareness) and then acting upon the collected information (teacher's interventions with the purpose of sustaining learning). LA provides a means for verifying or refuting assumptions a teacher has made about the CL classroom (Wise, 2018) and thus the withitness of the teacher might be increased with the help of LA. A dashboard incorporating AI in order to provide guidance might help teachers predict whether the students' collaboration is developing into a successful one or not. As a result, the effectiveness of the interventions the teacher subsequently decides to make (i.e., decides who needs support the most and takes supportive action) might also increase (Verbert et al., 2014). However, this supposition has not yet been confirmed by studies (Amarasinghe et al., 2020), which is why it is important to study how teachers make use of dashboards and other LA/AI devices in *authentic* settings, to determine whether such dashboards will indeed increase teachers' classroom awareness (van Leeuwen & Rummel, 2020). *We hypothesise that the introduction of a mirroring dashboard will increase the teacher's situational awareness in authentic settings (H1).*

**Table 1** Collaboration quality can be assessed using seven dimensions mapped with indicators for the teacher

Collaboration dimension	Theory	Indicators for teacher
Knowledge exchange	(Fischer & Mandl, 2003)	Are the students sharing information and clarifying each other's knowledge base? Are the students using each other as resources?
Sustaining mutual understanding	(Clark, 1996; Stahl, 2005; Thalemann & Strube, 2004)	Are the concepts used understood in the same way by the students? Are the students verbalising whether or not they understood one another, providing explicit feedback strategies?
Argumentation	(Tindale et al., 2003)	Are the students critically evaluating the information they have been provided with, which the group members bring to the table? Is the group comparing their outcome against the prompt/criteria?
Collaboration flow	(Clark, 1996)	Are the students making sure each member's attention is on the task? Are the students taking turns and coordinating the process of discussion?
Structure/Time	(Erkens et al., 2005)	Are the students planning their activities? Have they divided into roles and provided time for independent as well as joint discussion time?
Cooperative orientation	(Dillenbourg, 1999)	Are the students using a polite and respectful tone when talking to one another? If negative emotions or conflicts occur in the group, how are they managed?
Individual task orientation	(Barron, 2000; Johnson & Johnson, 2017)	How committed is each student towards the task? Motivation could be managed by strengthening individual accountability, removing distractions, directing attention towards relevant information, or expressing positive expectations towards the product



**Fig. 1** The teacher's situational awareness and skills on how to handle classroom situations make up teacher withitness

### From situational awareness to teacher interventions

Situational awareness combines both the understanding of the unfolding events as they are happening as well as the ability to predict what is going to happen next and whether the teacher's intervention is necessary. Gawron (2019) provides a guide for choosing the most suitable measure for situational awareness. In the case where the awareness is individual, and it is possible to stop the activity for a while, the recommendation is to use the Situation Awareness Global Assessment Technique (SAGAT) questionnaire. SAGAT is a performance measure, which has proven content, empirical and predictive validity, and is thus one of the most well-known measures of situational awareness. SAGAT uses statements where the participants need to respond with "TRUE" or "FALSE", which can be analysed using Signal Detection Theory suitable for binary answers (Abdi, 2009). Signal Detection Theory provides an assessment of how well the participant is able to spot a signal (hit). It enables analysis of sensitivity (i.e., the relative amount of hits instead of false alarms) across different conditions (Stanislaw, 1999). Signal Detection Theory can also be used to estimate the "answer bias" of each participant (i.e. the overall tendency to answer that a signal has been perceived). Teachers' interventions with the purpose of creating or sustaining a good quality collaboration can be regarded as *coregulation* methods. The teacher can coregulate by offering feedback or directly adapting the activities of the students. Scaffolding of group processes, not content, is expected in a coregulation process, meaning that teachers should not only provide tips on how to answer the next question or do the next task, but to help think about the process of learning (Hadwin & Oshige, 2011). Teachers need to manage their own teaching process while paying attention to the strategies students employ for solving the task and for the collaboration (van Leeuwen & Janssen, 2019). Nevertheless, in many real classroom conditions, it is not possible for one teacher to make sure each student is individually coregulated (Allal, 2020), which is where an alerting &

guiding dashboard might be of help (as opposed to a mirroring dashboard that only provides data about the students without any suggestions for intervention).

### **Using an alerting & guiding dashboard to decrease the workload of teachers**

Teacher-facing CSCL tools can broadly be grouped into three types: mirroring, alerting, and guiding tools (Soller et al., 2005). A mirroring LA tool only provides the teacher with some information about the collaboration without comparing it to a desired model (alerting) or offering suggestions for intervention based on quality predictions (a guiding tool). A systematic review of teacher guidance during CL (van Leeuwen & Janssen, 2019) revealed that the majority of the reported tools had only a mirroring function (as opposed to alerting or guiding). This means that some information is made accessible to the teacher, but the interpretation (explaining, predicting) and potential intervention thereof is currently left to the teacher, possibly increasing the workload of the teacher. This is a difficult task and demands a lot from the teacher, to see how challenging a task or situation is for a person, researchers can use the NASA- Task Load Index (TLX) instrument (Hart, 2006). Research on studying workload has shown that the mental demand decreases with the help of technology (Stadler et al., 2016), but participants with higher expertise have shown higher perceived effort and frustration levels (Kain et al., 2017). In addition to this, dashboard studies have mostly been carried out in lab settings (van Leeuwen et al., 2019), which help to understand how teachers could possibly use the data, but whether the findings will transfer to authentic settings remains debatable. As a notable exception, Kaliisa and Dolonen (2022) studied the use of a teacher-facing mirroring dashboard in authentic settings and accentuated the call for a guiding dashboard. A guiding dashboard based on an AI system predicting the future quality of the group could possibly aid the teacher in their decision making process. The teachers participating in their study using a mirroring dashboard wished for a dashboard integrated with AI. *We thus hypothesise that the introduction of a guiding dashboard will reduce the workload of the teacher in authentic settings (H2).*

### **Using an alerting & guiding dashboard to increase the coregulation skills of teachers**

Right now, the tools available may not support the teachers' reflective practice, because teachers are not always able to interpret the data presented to them and thus do not know how to respond or intervene (Dazo et al., 2017). From the perspective of the teacher's withitness, classroom awareness is supported by mirroring dashboards, but the decision of which intervention to choose (based on predictions constructed by the AI), is not supported. However, first results (Dickler et al., 2021) using guiding dashboards show that students' inquiry skills improved the most after the teacher supported student understanding and learning, instead of providing support in the subject content. Coregulation means that a teacher or more-capable peer is scaffolding a student to regulate their learning, while socially-shared regulation means that several people are regulating their common activity (Hadwin & Oshige, 2011). Shared regulation plays an integral role in successful collaboration (Järvelä et al., 2016) and research shows (Järvelä et al., 2019) that students are not capable of detecting demanding learning situations and their need for teacher coregulation on their own. Coregulation could be supported by providing the teacher with strategies for mirroring the students' ideas, requesting the students to reflect on their learning, modelling thinking, and offering prompts for thinking and reflecting. The teacher can coregulate by

offering feedback or directly adapting the activities of the students. Molenaar and Knoop-Van Campen (2018) report on teachers' feedback growing progressively more varied with the use of a dashboard. We thus hypothesize that *the introduction of an alerting & guiding dashboard increases teacher coregulation in authentic settings more than the introduction of a mirroring only dashboard (H3)*. Most dashboard studies have been conducted in more experimental settings and not in authentic classrooms, and if they were, the actual decision making and interventions by teachers were not studied.

## Methodology

The aim of this study was to explore how a mirroring only and an alerting & guiding dashboard affect teachers' management of a CL situation. To this end, based on the previous cycles of the design-based research (Kasepalu et al., 2022), three hypotheses have been formed:

H1: The introduction of a mirroring dashboard increases the teacher's situational awareness in authentic settings.

H2: The introduction of an alerting & guiding dashboard reduces the workload of the teacher in authentic settings more than the introduction of a mirroring only dashboard.

H3: The introduction of an alerting & guiding dashboard increases teacher coregulation in authentic settings more than the introduction of a mirroring only dashboard.

The withitness of the teacher comprises the situational awareness of the teacher (H1) and the interventions the teacher carries out in the classroom (H3). The workload of the teacher is operationalized as a measure of how difficult it is for the teacher to decide which intervention to choose based on the gained situational awareness (H2). We decided to carry out our study in authentic settings to see how the teacher would manage to understand the collaboration behaviour of the students and what kind of coregulation methods would be used in the field, if any. The study was designed as a quasi-experiment with a pre-test & post-test design, and with the application of a mirroring dashboard and an alerting & guiding dashboard as between-subject variables. The three independent variables are thus 1) no dashboard (control setting), 2) mirroring only dashboard, 3) alerting & guiding dashboard. Situational awareness (H1), workload (H2) as well as the types of classroom interventions (H3) were recorded as dependent measures (see Table 2 for all scales and measures).

## Participants

All participating teachers ( $N=24$ ) were in-service teachers with varying levels of work experience (range = 1–35 years of experience teaching, mean = 15.62 years,  $SD=11.09$ ). The teachers participated in multiple conditions. Of the teachers who volunteered to participate in the study, all twenty-four teachers participated in the no dashboard condition ( $N=24$ ). Nineteen of those also participated in the mirroring only dashboard condition ( $N=19$ ), and twelve teachers ( $N=12$ ) participated in the alerting & guiding dashboard condition. It was impossible for all teachers to participate in all three conditions due to technical difficulties (see Table 3 for the distribution of teachers). Ethics approval was obtained from the ethics board of the CEITER project and the ethics committee of Tallinn University (decision number 27). There were twenty-four teachers taking part in the



**Table 2** Concepts, scales, and measures introduced for each research question

	H1	H2	H3
Concept	situational awareness	workload	pedagogical actions: coregulation methods used
Method for data collection	self-report instrument SAGAT (binary scale) confidence rating (scale 1–4)	self-report instrument Raw Task Load Index (RTLX) (scale 1–100)	researcher’s classroom observations based on the Situated Model of Instructional Decision- Making
Data analysis	signal detection theory: hits, false alarms, response bias, sensitivity	unpaired T Test	number of actions carried out deductively, coded using the Situated Model of Instruc- tional Decision-Making

**Table 3** The participating teachers have been divided into three groups: 1) teachers involved in no dashboard and mirroring only dashboard condition ( $N=12$ ), 2) teachers involved in the pre-test and alerting & guiding dashboard condition ( $N=5$ ), 3) teachers involved in all three conditions ( $N=7$ )

Participating teachers	Conditions			N in the group
	No dashboard	Mirroring only dashboard	Alerting & guiding dashboard	
<b>Group 1</b>	x	x		12
<b>Group 2</b>	x		x	5
<b>Group 3</b>	x	x	x	7
N per condition	24	19	12	24

study from six schools over Estonia, of which twenty-two teachers were female and two were male. The teachers were teaching at either a middle school, upper-secondary school, or vocational school, and the data was collected between September 2021 and April 2022. All participating teachers and students gave informed consent. In the case of underage students, the consent from parents or caretakers was sought.

## Procedure

Prior to using the dashboards in lessons, the participating teachers answered a questionnaire inquiring about their overall teaching experience, knowledge, and confidence using collaborative learning methods. After this, together with the first author, a CL task was designed by the teacher using the collaboration analytics tool CoTrack (Chejara et al., 2022). The teachers were provided with a structure of a learning design but had to choose the topic, questions, and exact procedures according to the subject, class context and their personal preference. The teachers were briefed on the main functions of the dashboard prior to the learning design.

*Collaboration task design.* The students were asked to discuss an issue in subgroups (see Fig. 2 for an example classroom setup) for which leading questions were provided that the teacher formulated together with the first author. Subsequently, the students had to collectively construct a written discursive essay on the topic or carry out another collaborative

**Fig. 2** The classroom setup included each student using a personal computer with a microphone in a face-to-face setting



task compiled by the teacher. Each student had a personal computer to use and the possibility to modify a shared document using Etherpad (Kasepalu et al., 2022). The students and teacher were all in a face-to-face setting. Overall, classroom size varied from 12 to 30 students per lesson, which meant one teacher teaching three to eight groups at a time. The students were asked to sit in groups of four that either the students or the teacher had chosen for them, and each lesson lasted from 45 to 90 min. Altogether twenty lessons with eleven different class formations were run, meaning that some students took part in the study several times.

On the day of the data collection, the students were asked about what effective collaboration means to them and were introduced the seven dimensions of collaboration quality assessment. The students gave informed consent and could ask questions about the study and the platform. The data collection randomised the order of the three conditions that the teachers were presented with. This strategy was intended to avoid biasing the results due to teacher learning. The three conditions were: being an instructor of a CL task a) without a dashboard, b) with a mirroring only MMLA dashboard, c) with a mirroring MMLA dashboard together with an alerting and guiding dashboard paper prototype (see Table 4 for an overview of the three conditions).

During each condition, the first author observed the teacher and took note of her/his interventions. The teacher was free to operate the dashboard using a laptop whenever she/he wished in the b) and c) condition. After each condition, the participating teacher responded to SAGAT statements and the Raw Task Load Index (RTLX). After the lesson, the teacher took part in a semi-structured interview, where they were asked about the overall experience of using the dashboards and the reasons for deciding or deciding against intervening during the activity.

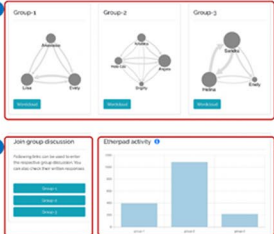

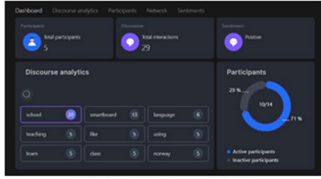
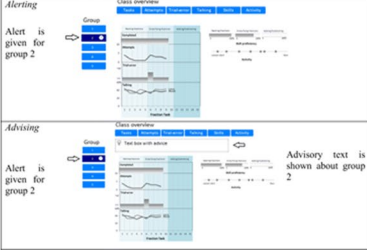
## Instruments and tools

We used a web-based tool CoTrack in our study. This tool offers functionality of collaborative writing using a real-time text editor, Etherpad. Additionally, it also has integrated data collection and processing features allowing generation of the dashboard in real-time. CoTrack collects audio data along with logs generated by collaborative writing. It then processes audio data in real-time and generates features derived from Voice Activity Detection, Speech-to-Text, Speaking time and Turn taking. From writing logs, it computes a student's contribution in terms of characters written or deleted in each collaborating group. These features are then used for generating two versions of the dashboard (see Fig. 3, explained in detail later): a mirroring only dashboard and an alerting & guiding dashboard. The mirroring dashboard is *automatically* generated by the tool and offers a *real-time* visualisation of group-level speaking and writing behaviour. In contrast, the alerting & guiding dashboard is generated in a combined manner by the tool and a researcher. The tool provides predicted levels of collaboration quality dimensions, and the researcher shows suggested coregulation strategies on paper (see Table 7 for the suggestions) to the teacher based on predicted results.

## Mirroring dashboard

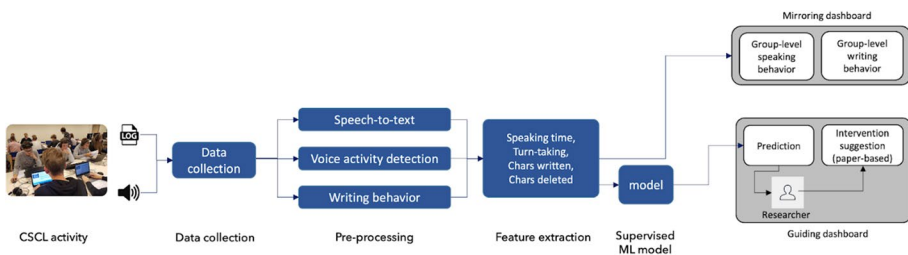
The mirroring dashboard visualises information about a collaborating group's speaking and writing behaviour. These features offer insight about an individual's participation in the group activity, which has been found as one of key quantitative metrics for CL (Weinberger

**Table 4** The three conditions applied in the study explained: in condition one the teachers did not use any dashboard, in condition two they were able to peruse a mirroring dashboard, whereas in the third condition they could interact with the mirroring dashboard, but were also given alerts and guidance from the dashboard

Condition 1: no support	Condition 2: mirroring only dashboard	Condition 3: alerting & guiding dashboard
The teacher carrying out CSCL in an authentic classroom without any external assistance	The teacher in an authentic CSCL scenario using dashboards that inform the teacher of the group progress and individual contribution of the students	The teacher in an authentic CSCL scenario using a system which also alerts the teacher of problematic groups and offers suggestions on how to possibly intervene during the CL activity to improve the collaboration process of the students
		
	 <p>(Kaliisa &amp; Dolonen, 2022)</p>	 <p>(van Leeuwen &amp; Rummel, 2019)</p>

& Fischer, 2006). The used features (see Table 5 for an overview of data features used) (e.g., speaking time, turn-taking) have also been found to be a good predictor of collaboration behaviour (Martinez et al., 2011).

The speaking behaviour is represented by the ‘who is talking after whom’ network (see number 2 in Fig. 4). The network consists of nodes and directed edges. The nodes represent the group’s members, and the edges indicate the ordering of speaking e.g., if participant A spoke after B then there will be an edge from node A to node B. The frequency of speaking



**Fig. 3** The anonymised tool and its generated dashboards

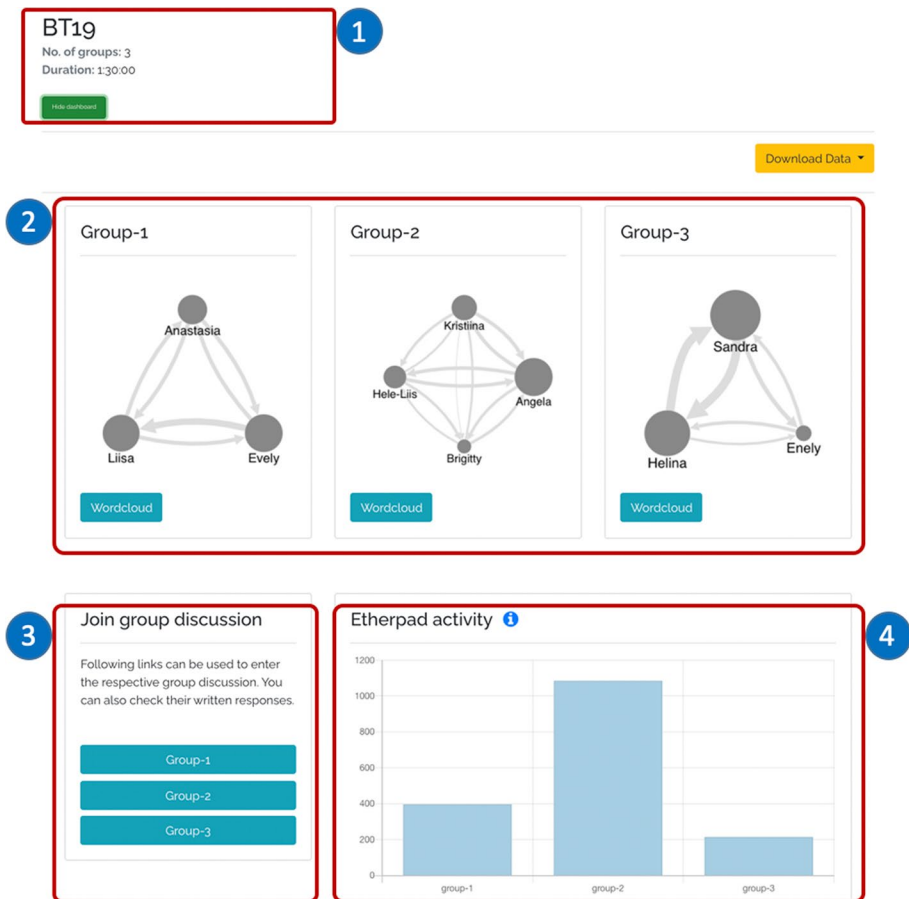
**Table 5** Data features used for developing the mirroring dashboard for CoTrack

Data source	Feature name	Description	Related studies
Audio	speaking_time	Speaking time in seconds	MMLA research studies have found speaking time and turn taking as indicators of collaboration (Ponce-López et al., 2015)
	turn_taking	Number of speaking turns taken by the participant	
	freq_I, freq_you & freq_we	Frequency of 'I', 'You', and 'We'	Stroch (2001) found differences between high and low performing collaborating groups in terms of their use of pronouns (e.g., I, We)
	freq_wh	Frequency of wh-words	Suggesting question-asking behaviour, which can reflect the argumentative nature of collaboration as per rating scheme of (Rummel et al., 2011)
Logs	chars_add	Number of characters added	An indicator of individual participation, which has been found as one of the key quantitative metrics for collaborative learning (Weinberger & Fischer, 2006)

in sequence in the group is encoded as the thickness of these edges. The more frequently participants A and B talk after each other, the thicker the edge between node A and node B will be. The dashboard also provides spoken text in the form of a word-cloud. The writing behaviour of every group is represented in the form of a bar graph (number 4 in Fig. 4). This graph shows the number of updates made by each group to their respective collaboratively written document. Additionally, the tool also allows the teacher to see the history of each update made in the document using a timeline.

### Alerting & guiding dashboard

The development of the alerting and guiding dashboard involved using a supervised machine learning technique for sub-dimensions of collaboration quality. We first annotated the ground truth of collaboration quality using the group's video recordings. The



**Fig. 4** Dashboard for collaboration activities using Jitsi and Etherpad (1—details for collaboration activity, e.g., title, duration, number of group, 2—group dynamics in terms of ‘who is talking after whom and how much’ and a button to show the word-cloud of the group, 3—controls for joining the group or to check group’s text written in Etherpad, 4—graphs showing the number of revisions made by each group in the Etherpad)

annotation process involved four graduate master students from The School of Digital Technologies. The training was provided in three rounds to all of them using a rating handbook from Rummel et al. (2011). Scores in the range of [-2, 2] were assigned for every dimension in the 30 s time window. The inter-rater reliability scores (Cohen's Kappa) were above 0.60 for each dimension of collaboration quality, indicating a substantial agreement (as per Landis and Koch (1977) guidelines).

For development of the model, we employed an image-based modelling approach. This approach was motivated by recent research on modelling collaboration in MMLA and found to achieve better performance (Anirudh & Dhinakaran, 2021). We represented our multimodal features (e.g., speaking time, turn-taking, writing operation, etc.) in the form of an image for every 30 s time window. The image represented features of each student in the group in a sequential manner, thus retaining the temporal order of students' different actions (e.g., speaking, writing, deleting). This representation enabled encoding of temporal information in addition to information from different audio and log features. We used a Convolutional Neural Network (CNN) with the images as data to train automated models for assessment of collaboration quality and its dimensions. We employed tenfold cross-validation for model evaluation. In this evaluation, the dataset was divided into 10 approximately equal parts; using one part for testing while the others for training in a repeated manner, where in each iteration a different part is selected for testing. Table 6 shows the performance of the developed models using three metrics: accuracy, area under curve (AUC), and Kappa.

For implementation of the current study, we have only employed those model dimensions to give feedback to teachers where there was suitable performance of the model. In addition to this, we took into account the comments of the teachers being interested in the individual contribution of each student, the atmosphere in the groups, and whether the students were feeling responsible for each other's understanding of content/ideas. Consequently, we chose individual task orientation, cooperative orientation, and sustaining mutual understanding for our investigation. Through machine learning,

**Table 6.** The performance of the developed models using three metrics: accuracy, area under curve (AUC), and Kappa, the chosen dimensions used for the guiding dashboard have been depicted in bold

Target	Accuracy	AUC	Kappa
Overall collaboration quality (CQ)	79%	87%	.45
<b>Sustaining mutual understanding (SMU)</b>	<b>80%</b>	<b>88%</b>	<b>.50</b>
Argumentation (ARG)	67%	72%	.33
<b>Cooperative orientation (CO)</b>	<b>78%</b>	<b>83%</b>	<b>.42</b>
Knowledge exchange (KE)	79%	85%	.42
Collaboration flow (CF)	76%	77%	.30
<b>Independent task orientation (ITO)</b>	<b>87%</b>	<b>88%</b>	<b>.38</b>
Structure and time planning (STR)	78%	85%	.43

the model has attempted to learn the association between several low level features and higher level constructs as suggested by the ACARS model (Rummel et al., 2011).

A similar approach is taken by Chen et al. (2017) to differentiate between productive threads and those needing improvement in collaborative knowledge building. Moreover, it should be noted that in contrast to Chen et al. and others who draw on text analytics, we draw on multimodal data, which means that the data features derive from several modalities (speech and text). In such a setting, it is quite common to construct more complex mappings between higher order constructs and data traces using machine learning (see e.g. Spikol et al. (2017).

To offer ideas for coregulation, we employed the Collaboration Intervention Model (CIM), which comprises coregulation strategies drawn from a detailed analysis of the literature (refer to (Kasepalu et al., 2022) for more details). CIM specifies which coregulation strategy could be offered on detection of ‘Low’ or ‘High’ state of a particular dimension of collaboration quality. For example, Table 7 shows a decision tree based on CIM for the individual task orientation dimension.

The dashboard alerted the teacher of the dimension that was low, thus helping the teacher interpret the data, and also guided the teacher with a hint for a possible coregulation suggestion (see Fig. 5, showing the reader one of the suggestions the teacher was given during the quasi-experiment where the dimension that was low was mentioned along with a suggestion for coregulation). The teachers were presented the name of the low dimensions and guidance in their native language.

## Adapted-SAGAT

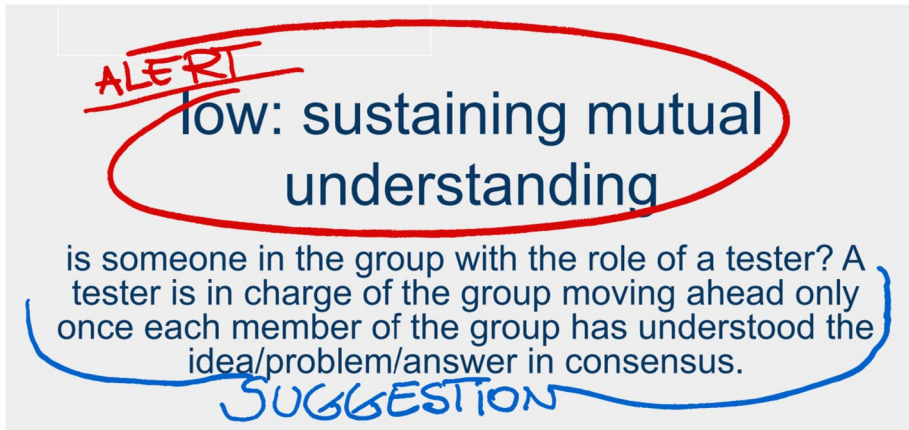
When studying individuals in a situation where it is possible to pause, the SAGAT methodology is advised (Gawron, 2019) for collecting data about situational awareness. In our case, we did not pause the lesson, the students carried on with their collaboration, but the teacher was given the questionnaire during the activity. Validated SAGAT statements (Endsley, 1995) with confidence ratings (Edgar et al., 2018) were used to collect information about the situational awareness of the teachers. First, 15 statements about CL situations were verbalised using the seven dimensions of the adaptable rating scheme modelled by Rummel et al. (2011) about three levels of situational awareness: perception, comprehension, and prediction (Endsley, 1995). Perception is characterised by the teacher monitoring and observing the classroom and students. The second level of situational awareness is comprehension, which looks at how the teacher uses the collected observations to understand what is happening around him/her. The third level of situational awareness is prediction, where the teacher is prognosticating what the outcome of the collaboration will be, how the group work will evolve, etc. The remaining 12 statements (see Appendix) were modified and corrected according to the advice of the evaluators and resulted in twelve pilot statements that were tested with seven in-service teachers.

Our goal was to carry out a study in an authentic setting. From the beginning, our work was set in an authentic classroom where teachers were already using a mirroring collaboration analytics tool to support CL activities, in which they were already responding to the twelve statements twice during a lesson. It took the teachers a lot of time to read and answer the questions. The time ranged from five to twenty-one minutes. The teachers were supposed to respond to the statements at least two times during a CL activity, both answering each question and also providing a confidence level rating per statement. The teachers in the pilot (as well as in the planned study) had to do this while simultaneously teaching



**Table 7** A fragment from CIM: intervention suggestions for individual task orientation dimension (see (Kasepalu et al., 2022) for the whole model)

Individual motivation (IND)	One student low	Doing nothing	<p>Build individual accountability into the task assessment (Barkley et al., 2014)</p> <p>Go and talk to the student, solicit the problem and guide the student in solving the problem (Barkley et al., 2014)</p> <p>Assign challenging roles for high achievers (Barkley et al., 2014)</p> <p>Alternative assignment (Barkley et al., 2014)</p> <p>Frontal instruction: have a thorough orientation in the benefits of collaborative learning (Barkley et al., 2014)</p> <p>Frontal instruction: open the floor for questions, when needed, provide explanations (Race, 2000)</p> <p>Do the students understand what these skills/concepts are used for? visualize/verbalize the connection with the course material/reality (Race, 2000)</p> <p>Go and see whether the students have understood the task; tell students exactly what they have to do and describe what the final product of their group work will look like (Johnson &amp; Johnson, 2017)</p>
	Several groups around 0 or less	Lack of motivation	
		Confusion	



**Fig. 5** The alerting and guiding features were displayed to the teacher on a piece of paper, thus alerting the teacher about which dimension of collaboration quality the group was low in (marked in red) and a suggestion guiding the teacher regarding how he/she could help the students in the particular group increase the quality of their collaboration (marked blue)

within the CL activities, which resulted in the teachers in the pilot deeming this approach as being too obtrusive. To shorten the questionnaire, we compared the means and standard deviations of the statements (without data and with a mirroring only dashboard) in order to identify and then eliminate those that had very low variability or no change. After consulting with two teachers in the pilot, the final four statements used in the data collection can be seen in here (but translated into Estonian for the study):

Perception

1. All students expressed their ideas in the group.

Understanding

2. All students were actively discussing with each other.

Prediction

3. All groups will finish the task in time.
4. All students will have learned from the task.

### Raw task load index

The NASA TLX is a measure of the difficulty of carrying out a task by a human. NASA-TLX uses six subscales: mental, physical, temporal demand, performance, level of effort and frustration. The first subscale is the mental demand of the teacher while measuring how much thinking, deciding, etc. was involved during the activity, whereas physical demand is the extent to which pulling, tugging, etc. was required from the teacher to teach in this environment. Thirdly, the teachers were asked about the temporal demands, asking how much time pressure the teacher felt during teaching. The fourth subscale,

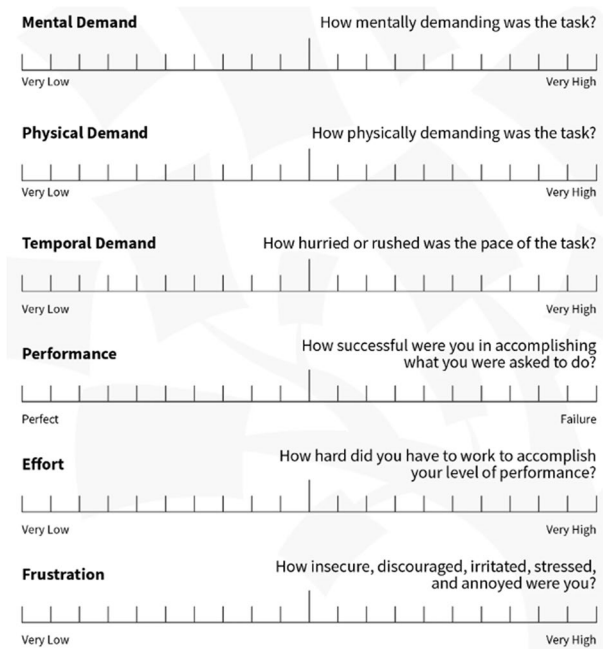
named performance, is an evaluation of how much of a failure the teacher met with while doing the activity. Next, the teachers were asked about the level of effort they had to put into teaching, aiming to find out how hard the teacher had to work. Lastly, the teachers were requested to reveal how insecure or discouraged they felt during CSCL, which is referred to as level of frustration. All the subscales are on a range between 0–100 (see Fig. 6 for the scale) meaning that the more mental demand the teacher perceived her/himself to have, the higher the number on the scale would be chosen.

Most commonly has this instrument been used to investigate the usability of an interface, especially in aviation, but computer and portable technology users have also been studied using NASA-TLX (Hart, 2006). The reported reliability is high: Cronbach's  $\alpha$  0.75 (Longo, 2018). We used the RTLX version due to time constraints, which means that all six dimensions are weighed the same instead of asking each individual which subscale ratings seem to affect their individual workload more. In addition to the RTLX, we also analysed the individual subscale ratings, as NASA-TLX offers a diagnostic value of the component subscales as well as the summative workload (Hart, 2006). As the study was carried out in an authentic classroom setting, it was important to be as unobtrusive as possible during the lessons.

## Data analysis

### Quantitative

RStudio was used to carry out the quantitative part of our data analysis. To study situational awareness, the number of hits (the ground truth and the teacher both say the adapted



**Fig. 6** RTLX scale 0–100. A higher number of e.g. physical demand indicates a higher workload on this subscale

SAGAT statement is true), the false alarms (the teacher says yes, but the ground truth is in reality a no), the misses (the ground truth reports yes, but the teacher misses it) were calculated (Stanislaw, 1999). The ground truth was established by the researcher using video analysis. In addition to this, we also report response bias (the general inclination of the participants to answer yes or no to the adapted SAGAT statements) and sensitivity (how good the teachers were at providing the correct answer, i.e., getting hits). A boxplot was used for visualisation purposes. To study the workload of the teachers, descriptive statistics methods were provided, the effect size (Cohen's  $d$ ) was calculated, and unpaired T-Tests were used due to not all teachers taking part in all three conditions.

## Qualitative

The first author took notes during the interventions carried out in the classroom. The coding scheme used for H3 was taken from the Situated Model of Instructional Decision-Making (Wise & Jung, 2019) where the interventions of teachers were coded into three main pedagogical actions: targeted action (which was separated into whole-class scaffolding and targeted group scaffolding), wait-and-see, and reflection. For the SAGAT statements, a researcher was present in all lessons and made direct observations of the students. During data analysis, statement 4 (about students having learned from the task) was discarded due to not having a ground truth to compare it to. For statements 1 and 2, the transcripts of the students' speech with the observations were used to establish the ground truth. For statement 3, the researcher examined the group product and checked whether it had been finished by the end of the lesson or not.

To illustrate the way the teachers used the dashboard, a narrative of the behaviour of T17 follows: *The English lesson starts, and the teacher seems excited to use the dashboard with the new guiding feature. This is the second time T17 is using the dashboard and she is already waiting for the word cloud function to start working. As this is the alerting & guiding dashboard condition for her, she can operate the mirroring dashboard at all times, and she follows all groups diligently throughout the lesson. She stops interacting with the mirroring dashboard for only about 10 min from the overall 90 min lesson. The researcher gave her 5 alerts and suggestions on how to proceed with two groups, but she just nodded, took notes, and carried on inspecting the written work of the students. There was no direct interaction with the students and the researcher only took note of wait-and-see and reflection actions in the classroom. The teacher later said that the dashboard is useful and she will use the information for student feedback.*

## Results

### **H1: The introduction of a mirroring only dashboard increases teachers' situational awareness.**

We hypothesised, based on previous studies, that the situational awareness of the teacher would increase with the introduction of a mirroring dashboard. Table 8 shows that the number of hits (the teacher provides the right answer= YES) in both dashboard conditions increased (with an additional advantage for the alerting & guiding dashboard) as compared to the no dashboard condition. However, the sensitivity of the teachers increased with the introduction of the mirroring dashboard from -0.05 to 1.9.

**Table 8** The signal detection theory average features of the teachers across the three conditions of our study

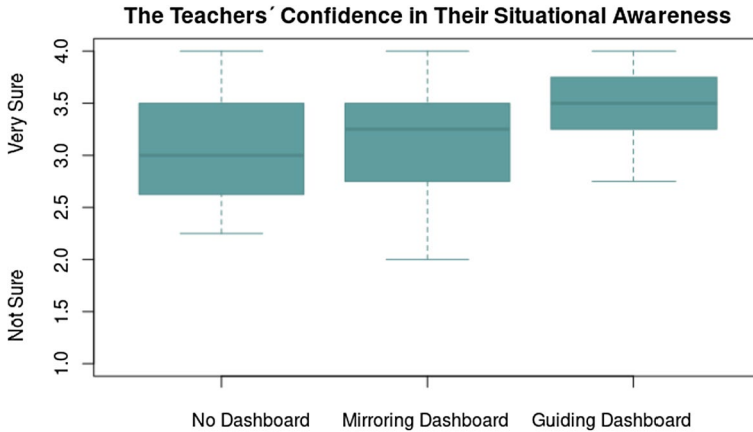
Situational awareness	No dashboard	Mirroring only dashboard	Alerting & guiding dashboard
hits	32.0%	69.0%	77.8%
false alarms	34.0%	9.0%	8.3%
sensitivity	-0.1	1.9	1.4
response bias	-18.8	-15.4	-8.5

This sensitivity score is computed as the amount of true positives, which occur when the signal is correctly detected. This means that the teachers detected more true positives in the mirroring dashboard condition than they did in the no dashboard condition. As an example, in our case the teacher correctly detected that the groups would finish the task in time. The sensitivity was also the highest in the mirroring dashboard condition. A one-way between subjects ANOVA was conducted to compare the effect of the condition on the sensitivity of the participants in these authentic settings. There was a significant effect of the condition on sensitivity at the  $p < 0.05$  level for the three conditions ( $F = 22.1$ ,  $p < 0.001$ ). The response bias (i.e., the likelihood to either answer YES or NO) is negative in our case, meaning that the teachers were more prone to answer “yes” to the statements. A between-subjects one-way ANOVA revealed that the condition has no statistically significant effect on the response bias of the teachers ( $F = 0.7$ ,  $p = 0.42$ ). All statements collected were consistent with the idea that the students were having an effective collaboration. So the teachers found that their students were mostly collaborating effectively, though this perception may not have been entirely accurate. To better understand the metric, a hypothetical person without any bias would score 0. However, in our study the lowest response bias can be discerned in the alerting & guiding dashboard condition (from  $-18.8$  to  $-8.5$ ), which possibly helps the teachers have a less biased understanding of what is happening in the classroom.

When the participating teachers had to decide whether the situational awareness statements were true or false for the situation at hand, they also had to indicate their level of confidence from 1 (just guessing) to 4 (very confident). In Fig. 7 we can see that the teachers feel the most confident in the alerting & guiding dashboard condition, but also that they are generally quite confident (we may be witnessing a ‘ceiling effect’).

## H2: The introduction of an alerting & guiding dashboard reduces the workload of teachers

As the workload measure we used RTLX, where all six subscales are weighted the same and the average per participant was calculated. The overall workload of the teachers participating in the quasi-experiment suggested that the alerting & guiding dashboard decreased the workload of the teacher using it (from 35.6 in the no dashboard condition to 24.4 in the alerting & guiding dashboard condition, see Table 9 below). An alerting & guiding dashboard reduced the workload of the teacher even more than the mirroring dashboard (30.4 vs 24.4 in the alerting & guiding dashboard condition). A between-subjects one-way ANOVA revealed that the dashboard condition has a significant effect on the overall workload of the teachers ( $F = 4.18$ ,  $p = 0.046$ ). We further compared the



**Fig. 7** The participating teachers are confident in their awareness

conditions using unpaired T-Tests to see where the biggest differences could be seen. Having compared the no dashboard and an alerting & guiding dashboard condition, a statistically significant difference ( $p=0.000003$ ) with a big effect (Cohen’s  $d=1.6$ ) (Cohen, 1998; Navarro, 2015) was detected in the overall workload.

Furthermore, when comparing the mirroring only and alerting & guiding dashboard conditions, the overall workload of the teachers decreased in the alerting & guiding dashboard condition, having a big effect (1.1) and being statistically significant ( $p=0.002$ ). This suggest that the introduction of an alerting & guiding dashboard seems to significantly decrease the workload of the teacher.

When looking at the subscales of the teacher workload, the performance and frustration subscales stood out as experiencing the sharpest drop. In addition to the overall workload, two subscales likewise demonstrated a difference with a large effect. These two subscales were performance (how accomplished and successful the teacher was feeling during the CL activity, Cohen’s  $d=1.2$ ,  $p=0.003$ ) and level of frustration (how discouraged, stressed and annoyed the teacher was feeling during the CL activity, Cohen’s  $d=1$ ,  $p=0.005$ ). This

**Table 9.** The average workload of the teachers under three conditions, the coloured types of workloads show the biggest difference between conditions

Type of load (SD)	No dashboard	Mirroring only dashboard	Alerting & guiding dashboard
overall workload	35.6 (14.5)	30.4 (18.5)	24.4 (14.1)
cognitive demand	38.1 (24.4)	31.3 (25.8)	32.5 (19.7)
physical demand	20.6 (18.3)	16.3 (20.7)	17.5 (19.7)
temporal demand	41.7 (24.7)	36.1 (31.4)	32.1 (21.6)
performance	44.0 (21.1)	37.1 (22.5)	20.8 (19.3)
effort	41.5 (20.1)	39.5 (24.0)	33.3 (26.8)
frustration level	28.1 (21.5)	22.4 (26.9)	10.0 (14.1)

implies that the teachers might feel their performance is increased and frustration decreased when teaching with the help of an alerting & guiding dashboard, when compared to the no dashboard condition. Differences in the other dimensions like cognitive load, temporal demand and effort were non-significant, even though a similar pattern in means was observed. Additionally, the teachers appeared to feel more accomplished (performance subscale) in the alerting & guiding dashboard condition compared to the mirroring only dashboard condition (medium effect 0.8,  $p=0.04$ ).

### **H3: The introduction of an alerting & guiding dashboard increases teacher coregulation in authentic settings more than the introduction of a mirroring only dashboard**

It can be perceived from Table 10 that the mirroring dashboard seemingly increased the number of teachers intervening and reduced the number of teachers who did not intervene or adopted the wait-and-see strategy. In the final column in Table 10 we have reported the expected equal distribution between the three different conditions: no dashboard, mirroring dashboard only, and alerting & guiding dashboard. This enables us to compare each condition's observed number of occurrences with the expected value. The mirroring condition seemed to activate the teachers into coregulating more, three teachers started a coregulation intervention frontally and two coregulation interventions were initiated in groups. It needs to be mentioned that teachers commenced coregulating the whole group of students solely in the mirroring only condition. For instance, T13 after having observed the students and the mirroring dashboard for a while, stopped the activity of the students and asked the students what the goal of the task was. T13 reported that the word clouds and rather much inactivity from the students had indicated to her/him that the students might not know exactly what to do. After a short discussion about the connection to previously studied material and a project they were going to work on later on, the students phrased the goal of the collaborative task in a way that was suitable for the teacher. What is more, three teachers out of twelve introduced a targeted coregulation within groups. This one out of four ratio is much higher than the ratio of teachers initiating in the first two conditions (1 out of 24 in the no dashboard condition and 2 out of 19 in the mirroring only condition).

Overall, the mirroring only condition activated the teachers into intervening more (the participating teachers carried out an intervention in the classroom in 45.8% (11 teachers out of 24) of the cases in the no dashboard condition to 73.7% (14 out of 19) in the mirroring only condition). There was a difference that proved non-significant under a Chi-squared test (5.62,  $p=0.06$ ). Nevertheless, T29 said that the mirroring dashboard “*showed me which students were being inactive and I could actually base my decision which group to approach on tangible evidence*”. However, it must be said that only a meagre portion of these interventions were coregulation strategies. Most often the teachers went up to the students and pointed at a mistake or guided them towards a place in their materials. In the alerting & guiding dashboard condition, however, only 33.3% of the participating teachers started an intervention during the CL activity, which means that we did not prove the third hypothesis. Nevertheless, three out of these interventions could be considered coregulation strategies as they were asking questions about the process, or providing prompts for activating members of the group, much more when compared to the no dashboard condition. We need to interject that three of the teachers intervening in the mirroring only condition had first gone through the alerting & guiding dashboard condition, meaning that they could have needed a

**Table 10** The interventions of the participating teachers coded using the situated model of instructional decision-making. The percentage in columns 2–4 is calculated from all participating teachers in the specific condition. Column 5, called Expected Equal Distribution, provides a percentage under the assumption that all intervening teachers had been divided equally among three conditions: no dashboard, mirroring only dashboard and alerting & guiding dashboard. Highlighted values vary from equal, proportionate division greatly

	No Dashboard	Mirroring Only Dashboard	Alerting & Guiding Dashboard	Expected Distribution in Case of Independence
<b>INTERVENTION</b>	45.8% (11/24)	73.7% (14/19)	33.3% (4/12)	52.7% (29/55)
Coregulation	4.2% (1/24)	26.3% (5/19)	25% (3/12)	16.4% (9/55)
Scaffolding	41.7% (10/24)	73.7% (14/19)	33.3% (4/12)	50.9% (28/55)
<b>NO INTERVENTION</b>	54.2% (13/24)	26.3% (5/19)	66.7% (8/12)	47.3% (26/55)
Wait and See	33.3% (8/24)	15.8% (3/19)	25% (3/12)	25.5% (14/55)
Reflection	-	5.3% (1/19)	33.3% (4/12)	9% (5/55)
Do Nothing	21% (5/24)	5.3% (1/19)	8.3% (1/12)	12.7% (7/55)

little time for reflection, and after internalising the strategy, started using it in the mirroring only condition. When we asked the non-intervening teachers why they had not intervened, four teachers commented that they would have done it, had the dashboard alerted the collaboration quality to be low for a *longer* period. T20 declared that she “*would only have coregulated the process if there had been no progress at all within the group*”. Four teachers said that although the suggestions provided by the alerting & guiding dashboard seemed useful, they needed some time to process them and think how to work them into the next CL activity initiating a reflection process more than in other conditions.

As we had collected data about the knowledge, practice and understanding of CL from the teachers, we also carried out comparative analyses to see whether some variables would explain the differences, but the differences were all marginal. Regarding the use for the alerting & guiding dashboard suggestions (which may explain the observed effects on propensity to introduce interventions), eight teachers said the alerting & guiding dashboard helped them identify students not participating and in making the decision whether an intervention was necessary. Six teachers said that the suggestions they had been given provided them with ideas on how to make the collaboration more effective, but not all of them carried out a coregulation strategy in the classroom. T17 saw the dashboard as a way of collecting information to prepare for providing individual feedback to the students. The



suggestions introduced by the alerting & guiding dashboard do not necessarily need to be brand new information, as T15 said that “*amid all the classroom turmoil, it was great when I got these suggestions as the mind can sometimes just go blank even when I am able to identify a group in need myself. This way it acts as sort of a buffer, an additional resource*”.

## Discussion

*Withitness of the teachers increased with the introduction of a mirroring only dashboard.* The situational awareness of the teachers increased with the help of a mirroring and alerting & guiding dashboard as had previously been shown in several controlled studies (van Leeuwen & Rummel, 2020; Verbert et al., 2014), but not previously studied in an authentic setting. However, when van Leeuwen et al. (2019) studied teacher detection (comparable to the perception phase in situational awareness), using a mirroring, alerting, and guiding dashboard in controlled settings, they found no statistically significant differences. In authentic settings, where the teachers know the students and they are interacting with them in real time, the situation for the teachers is more high stakes. Our results, collected in a particular authentic setting provide a very strong indication that introducing dashboards for teachers can improve teachers’ decision making in the classroom by being better at indicating problems (higher sensitivity) and having lower response bias. Nevertheless, more studies in authentic settings are needed to put the teachers in a situation closest to their real practice and to validate that our results are not context specific.

It could be perceived that *the response bias decreased with the introduction of an alerting & guiding dashboard*, meaning that the teachers grew less biased towards their awareness of the classroom situation with the help of an alerting & guiding dashboard. However, we could see that the teachers’ confidence may have had a ceiling effect, meaning that their confidence in their situational awareness was rather high to start with. Nonetheless, when we compare the results in the three conditions, it could be seen that although the teachers were rather confident in the no dashboard condition, 50% of the predictions made by the teachers without using a dashboard had been incorrect, pointing to an overconfidence bias. Teachers are used to making decisions in a situation with a lack of data and although the introduction of a mirroring only or alerting & guiding dashboard helps them perceive, comprehend, and predict the situation better, they mostly feel rather confident without it as well. In which case, if the dashboards would help to reduce this overconfidence, then this could be seen as a positive outcome in itself. In addition to having a ceiling effect, the teachers were biased towards thinking that their students are collaborating in a more effective way than they were actually doing. This points us to the need of training actions that complement the use of MMLA and LA tools. Training for teachers is needed in assessing collaboration as well as using tools to base this assessment on reliable data.

*There is strong evidence that dashboards reduce the workload of the teachers.* When van Leeuwen et al. (2019) studied only the cognitive load of teachers using different dashboards, they found that the introduction of a dashboard lowered the teachers’ workload without statistical significance. As we studied all six sub-categories of workload, we could see that two sub-categories, namely performance and level of frustration, had more effect on the overall workload rather than cognitive load specifically. We need to consider that LA and providing suggestions based on the AI quality assessment to a teacher is adding information to be processed in our implementation (in the alerting & guiding dashboard condition, the teachers could still have access to the mirroring dashboard). Taking this into account, it seems reasonable that the reduction in cognitive load is not as prevalent as the feeling of working in a more

effective and efficient way. The feeling of frustration and effort were associated with positive but non-significant differences, thus larger studies are needed to understand whether there is a real effect or not. As teachers in our sample were feeling less frustration and less like a failure when using a dashboard, dashboards might have an influence on the emotional state of the teacher. Further studies could investigate the emotional impact of these technologies, or how to make the human-AI interaction better for emotional impact specifically. An implication for further research is to study the workload with sub-dimensions with a larger sample of teachers.

*Teacher interventions during CL increase with a mirroring dashboard, whereas an alerting & guiding dashboard initiates teacher reflection.* Previous studies (Kaliisa & Dolonen, 2022; Kasepalu et al., 2021) have shown that teachers might need more scaffolding from the dashboard for the data to be actionable as low-level data may not provoke an intervention. Notwithstanding, after having conducted a quasi-experiment in authentic settings, we could see that the number of interactions did not increase in the alerting & guiding dashboard condition. However, we could perceive that more coregulation strategies were employed in the mirroring only and alerting & guiding dashboard conditions compared to the no dashboard condition. When teachers were prompted with coregulation strategies, then they were more likely to intervene using a coregulation strategy instead of just guiding the students towards the right page number in a book or correcting a mistake the students had made. It must be noted that three teachers started employing coregulation strategies after they had interacted with the alerting & guiding dashboard beforehand. Nevertheless, starting to use the presented suggestions based on AI quality evaluation could be difficult if the suggestion is novel to the teacher and she/he has no prior experience using it. Notwithstanding, several teachers claimed to have started a reflection process after an interaction with the AI assistant, getting data on the predicted quality of collaboration and the suggestions provided, the changes in teachers might have been much more nuanced than could be perceived within the experimental protocol in class or in a short interview, but would rather need insights from a reflective journal (Park & Zhang, 2022). They said that they needed some time to think and possibly use the suggestion in the next lesson or next group task. This is consistent with the previous results reported by van Leeuwen et al. (2019) who showed that the response time for teachers was higher in the alerting & guiding dashboard condition vs the mirroring only dashboard condition. Molenaar and Knoop-Van Campen (2018) showcased in their study that the teachers' pedagogical actions grew progressively more varied over time, but in our study the teachers had a limited time to interact with the dashboard. We measured and observed authentic lessons with the teachers in one or two lessons, where the teachers started to get acquainted with the tool and understand its possibilities. We suggest that it takes more training (e.g., teachers in the study Kaliisa and Dolonen (2022) suggest peer training). Training might occur during a workshop or a longer period of trying the tool to see an increase in interventions. The change should be especially in coregulation interventions focused on the process of collaboration as this is the type of support that has been confirmed to show the highest improvement of students' skills (Dickler et al., 2021). Collaboration skill is a general skill, which in the Estonian context no specific subject teacher is responsible for teaching. Could this be leading to a situation where every teacher should be guiding and coregulating the students, but in the end, no one is doing it? Could this be improved by updating policies, curricula or maybe carrying out sporadic assessment of student collaboration skills?

*The teacher needs to be an active agent using dashboards.* Similarly to studies carried out in controlled settings (van Leeuwen et al., 2019), the teachers did not blindly accept the suggestions of the dashboard. Right now, both the mirroring and alerting & guiding dashboard were rather rigid and not much configurable by the teacher, which might be decreasing teacher agency

(Kaliisa & Dolonen, 2022) and thus adding to the unwillingness to adopt the suggestions. An implication for study and dashboard design is to consider teacher agency as an important factor for adopting new tools. Similarly, to research conducted by Vieira et al. (2018) in our authentic setting, the participating teachers voiced the concern of having to deal with hardware, computer settings and internet connection problems, which could additionally restrain the use of such tools. With the intention of growing the theory-based Collaboration Intervention Model organically with practice-based expert-validated suggestions, we imagine providing the teacher with the authority to add “interventions that worked” and connecting them to dimensions of collaboration.

The limitations of the study include a small voluntary sample of teachers. Due to technical difficulties, not all interested teachers were able to engage in the study in all three conditions, which is why stronger (paired) statistical tests could not be conducted. This could possibly have introduced systematic bias into our results, the probability of which was tested by comparing the teachers in the two different experimental conditions. We found no big differences between the groups in terms of the studied variables. The classes were not evenly distributed, which is the reason why the accuracy metric was above 77%, whereas the kappa was in the range 0.38-0.50 for all three dimensions used in the study. The kappa was fair for ITO and moderate for SMU and CO as per guidelines from Landis and Koch (1977). However, it was still low, and we aim to improve it using additional features (e.g., speech features) for our future studies. Different confounding variables might have influenced the teachers in this authentic setting, e.g., the different students in the lessons, the teacher’s mental wellbeing etc. Furthermore, the different ways of presenting information (visual in the mirroring dashboard, whereas visual together with lexical in the alerting & guiding dashboard condition) could have affected the results. We are not fully able to discern whether the effects were due to supporting interpretation or guidance as the third condition involved both of the features. This study design was chosen due to our strong belief in explainable AI and that teacher actions should only be recommended by AI systems, if there is a reasoned and pedagogically grounded interpretation given. In addition to this, the multiple hypothesis testing problem needs to be considered and the fact that our study was conducted in a specific cultural context in Estonia. Studies in different national and cultural contexts need to be carried out in the future. The alerting & guiding dashboard was a paper prototype, the background analysis of which had not been explained to the teachers, which might have made it more difficult for teachers to trust it and therefore might have deterred them from implementing the provided suggestions.

## Conclusion and future work

Studying teacher withitness in the wild demonstrated that a mirroring only dashboard increased the teachers’ situational awareness, and an alerting & guiding dashboard integrating an AI assistant predicting the quality of collaboration decreased the teachers’ workload. The teachers intervened more in the mirroring only dashboard condition when compared to the no dashboard condition. Both dashboard conditions increased the number of coregulation interventions carried out, however, not to the extent that we had hypothesized. This research makes a contribution to the CSCL literature as well as the general classroom orchestration field. Based on our results, we suggest the following implications:

There is a necessity for *larger authentic studies*, the conducting of which is difficult, messy and has many confounding variables which are difficult to control. We suggest using four different conditions: no dashboard; mirroring only dashboard; mirroring and alerting dashboard; mirroring, alerting, and guiding dashboard. Notwithstanding the dire straits, observing teachers

interacting with LA in authentic settings gives a more insightful picture into the reality of a teacher using a teacher facing dashboard and the difficulties they confront. The teachers feel that the data is useful, but it is unclear whether and how they would use it in their everyday practice. Even though in hypothetical or controlled settings the system had suggested different interventions by the teachers, and them becoming active as a response to the stimuli of the dashboard, our results did not provide a confirmation. We suggest new variables to be added (experience, workload), and better tooling introduced (Chejara et al., 2022) to make it possible to learn from such complex authentic experiments how to aid teachers' withitness in the CL classroom.

As the results showcased that the reaction of the teachers varied highly in an authentic classroom, conducting longer-term case studies would help understand the decision-making process of teachers better. We propose *a longitudinal case study design observing teachers using the alerting & guiding dashboard in their lessons*, taking note of their interventions, having the teachers write reflective journal entries and having in depth interviews with the teacher afterwards to understand why certain suggestions were followed and others not.

*Teacher agency* needs to be considered in the design of dashboards. The question is how to make the data and analysis transparent to the teacher in a way that is understandable and does not overburden the teacher with information. The teachers, and especially more experienced teachers, need to understand which data and models the predictions are based on to trust the analysis and suggestions provided by the AI. For a teacher to have AI amplified intelligence, she/he needs to feel they are a partner to the AI, not a servant. Teacher agency needs to be studied when using teacher dashboards. We also suggest using teacher experience as an experimental variable, possibly having extreme groups, to see whether the perception and use of dashboards vary. In addition to this, based on the positive effect of the dashboards in our sample on the teacher's level of frustration and feeling a failure, we suggest studying the influence AI has on the emotional state of the teacher. Further studies could investigate the emotional impact of these technologies, or how to make the human-AI interaction better emotionally.

## Appendix

SAGAT statements for the piloting stage:

1. There is good communication between group members in all groups.
2. The students were motivated to do the task.
3. The students in the groups listened to one another.
4. The students expressed their ideas in the group.
5. All students were engaged in the task most of the time.
6. Students were actively discussing with each other.
7. All groups progressed at a similar rate.
8. In all groups the members clarified each other's ideas.
9. All groups will finish the task in time.
10. The task result of all groups will be of threshold level.
11. All students will have shared knowledge in the group.
12. All students enhanced their collaboration skills.

**Data Availability** Data available on request from the authors.

## Declarations

**Competing interests** No competing interests to report.

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

## References

- Abdi, H. (2009). Signal Detection Theory (SDT). In *Encyclopedia of education* (3rd ed., pp. 10). Elsevier. <https://doi.org/10.4135/9781412983907.n1799>
- Allal, L. (2020). Assessment and the co-regulation of learning in the classroom. *Assessment in Education: Principles, Policy & Practice*, 27(4), 332–349. <https://doi.org/10.1080/0969594x.2019.1609411>
- Amarasinghe, I., Hernandez-Leo, D., Michos, K., & Vujovic, M. (2020). An actionable orchestration dashboard to enhance collaboration in the classroom. *IEEE Transactions on Learning Technologies*, 13(4), 662–675. <https://doi.org/10.1109/TLT.2020.3028597>
- Anirudh, K., & Dhinakaran, S. (2021). Numerical analysis of the performance improvement of a flat-plate solar collector using conjugated porous blocks. *Renewable Energy*, 172, 382–391. <https://doi.org/10.1016/j.renene.2021.02.145>
- Baker, R. S. (2016). Stupid tutoring systems, intelligent humans. *International Journal of Artificial Intelligence in Education*, 26(2), 600–614. <https://doi.org/10.1007/s40593-016-0105-0>
- Barkley, F. E., Major, C. H., & Cross, K. P. (2014). *Collaborative learning techniques. A handbook for college faculty*. (Second Edn). Jossey-Bass. A Wiley Brand.
- Barron, B. (2000). Achieving coordination in collaborative problem-solving groups. *Journal of the Learning Sciences*, 9(4), 403–436. [https://doi.org/10.1207/S15327809JLS0904\\_2](https://doi.org/10.1207/S15327809JLS0904_2)
- Chejara, P., Kasepalu, R., Ruiz-Calleja, A., Rodríguez-Triana, M., & Prieto, P., L. (2022). Co-designing a multimodal dashboard for collaborative analytics. *The 15th International Conference on Computer-Supported Collaborative Learning (CSCL)*, 5535 LNCS, 331–336. [https://doi.org/10.1007/978-3-642-02247-0\\_32](https://doi.org/10.1007/978-3-642-02247-0_32)
- Chen, B., Resendes, M., Chai, C. S., & Hong, H. Y. (2017). Two tales of time: Uncovering the significance of sequential patterns among contribution types in knowledge-building discourse. *Interactive Learning Environments*, 25(2), 162–175. <https://doi.org/10.1080/10494820.2016.1276081>
- Clark, H. (1996). *Using Language* (6th ed.). Cambridge University Press.
- Cohen, J. (1998). *Statistical power analysis for the behavioral sciences* (2nd ed.).
- Dazo, S. L., Stepanek, N. R., Chauhan, A., & Dorn, B. (2017). Examining instructor use of learning analytics. *Conference on Human Factors in Computing Systems - Proceedings, Part, F1276*, 2504–2510. <https://doi.org/10.1145/3027063.3053256>
- Dickler, R., Gobert, J., & Pedro, M. S. (2021). Using innovative methods to explore the potential of an alerting dashboard for science inquiry. *Journal of Learning Analytics*, 8(2), 105–122. <https://doi.org/10.18608/JLA.2021.7153>
- Dillenbourg, P. (1999). Introduction: What do you mean by “collaborative learning”? *Collaborative Learning: Cognitive and Computational Approaches*, 1, 1–19.
- Dillenbourg, P. (2013). Design for classroom orchestration. *Computers & Education*, 69, 485–492.
- Edgar, G. K., Catherwood, D., Baker, S., Sallis, G., Bertels, M., Edgar, H. E., Nikolla, D., Buckle, S., Goodwin, C., & Whelan, A. (2018). Quantitative Analysis of Situation Awareness (QASA): Modelling and measuring situation awareness using signal detection theory. *Ergonomics*, 61(6), 762–777. <https://doi.org/10.1080/00140139.2017.1420238>

- Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37(1), 32–64. <https://doi.org/10.1518/001872095779049543>
- Erkens, G., Jaspers, J., Prangsmas, M., & Kanselaar, G. (2005). Coordination processes in computer supported collaborative writing. *Computers in Human Behavior*, 21(3 SPEC. ISS.), 463–486. <https://doi.org/10.1016/j.chb.2004.10.038>
- Fischer, F., & Mandl, H. (2003). Being there or being where? Videoconferencing and cooperative learning. In H. van Oostendorp (Ed.), *Cognition in a Digital World* (pp. 205–223). Lawrence Erlbaum Associates.
- Gawron, V. J. (2019). Measures of situational awareness. In *Human performance and situation awareness measures*. <https://doi.org/10.1201/9780429001024-3>
- Gegenfurtner, A., Lewalter, D., Lehtinen, E., Schmidt, M., & Gruber, H. (2020). Teacher expertise and professional vision: Examining knowledge-based reasoning of pre-service teachers, in-service teachers, and school principals. *Frontiers in Education*, 5(May), 1–8. <https://doi.org/10.3389/educ.2020.00059>
- Hadwin, A., & Oshige, M. (2011). Self-regulation, coregulation, and socially shared regulation: Exploring perspectives of social in self-regulated learning theory. *Teachers College Record*, 113(2), 240–264.
- Hart, S. G. (2006). *Nasa-task load index (NASA-TLX); 20 years later*. 904–908.
- Holstein, K., McLaren, B. M., & Aleven, V. (2019a). Co-designing a real-time classroom orchestration tool to support teacher–AI complementarity. *Journal of Learning Analytics*, 6(2), 27–52. <https://doi.org/10.18608/jla.2019.62.3>
- Holstein, K., McLaren, B. M., & Aleven, V. (2019b). Designing for complementarity: Teacher and student needs for orchestration support in AI-enhanced classrooms. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*: Vol. 11625 LNAI. Springer International Publishing. [https://doi.org/10.1007/978-3-030-23204-7\\_14](https://doi.org/10.1007/978-3-030-23204-7_14)
- Järvelä, S., Järvenoja, H., & Malmberg, J. (2019). Capturing the dynamic and cyclical nature of regulation: Methodological Progress in understanding socially shared regulation in learning. *International Journal of Computer-Supported Collaborative Learning*, 14(4), 425–441. <https://doi.org/10.1007/s11412-019-09313-2>
- Järvelä, S., Malmberg, J., & Koivuniemi, M. (2016). Recognizing socially shared regulation by using the temporal sequences of online chat and logs in CSCL. *Learning and Instruction*, 42, 1–11. <https://doi.org/10.1016/j.learninstruc.2015.10.006>
- Johnson, D., & Johnson, R. (2017). Cooperative learning. In G. de Aragon (Ed.), *Innovación educación. I Congreso Internacional*. (pp. 12). Gobierno de Aragon. file:///C:/Users/lenovo/Downloads/David\_Johnson (3).pdf.
- Kain, K. S., Stadler, S., Giuliani, M., Mirnig, N., Stollnberger, G., & Tscheligi, M. (2017). *Tablet-based augmented reality in the factory*. 151–152. <https://doi.org/10.1145/3029798.3038347>
- Kaliisa, R., & Dolonen, J. A. (2022). CADA: A teacher-facing learning analytics dashboard to foster teachers' awareness of students' participation and discourse patterns in online discussions. *Technology, Knowledge and Learning*. <https://doi.org/10.1007/s10758-022-09598-7>
- Kasepalu, R., Prieto, L. P., Ley, T., & Chejara, P. (2022). Teacher artificial intelligence-supported pedagogical actions in collaborative learning coregulation: A Wizard-of-Oz Study. *Frontiers in Education*, 7. <https://doi.org/10.3389/educ.2022.736194>
- Kasepalu, R., Chejara, P., Prieto, L. P., & Ley, T. (2021). Do teachers find dashboards trustworthy, actionable and useful? A vignette study using a logs and audio. *Technology, Knowledge and Learning*, 0123456789, 1–19. <https://doi.org/10.1007/s10758-021-09522-5>
- Kobbe, L., Weinberger, A., Dillenbourg, P., Harrer, A., Hämmäläinen, R., Häkkinen, P., & Fischer, F. (2007). Specifying computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning*, 2(2–3), 211–224. <https://doi.org/10.1007/s11412-007-9014-4>
- Kounin, J. (1970). *Discipline and group management in classrooms*. Rinehart and Winston.
- Landis, J. R., & Koch, G. G. (1977). *The measurement of observer agreement for categorical data* (Vol. 33, Issue 1). <https://www.jstor.org/stable/2529310>. Accessed 6 Dec 2023.
- Longo, L. (2018). On the reliability, validity and sensitivity of three mental workload assessment techniques for the evaluation of instructional designs: A case study in a third-level course. *CSEDU 2018 - Proceedings of the 10th International Conference on Computer Supported Education*, 2, 166–178. <https://doi.org/10.5220/0006801801660178>
- Martinez, R., Kay, J., Wallace, J. R., & Yacef, K. (2011). Modelling symmetry of activity as an indicator of collocated group collaboration. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6787 LNCS, 207–218. [https://doi.org/10.1007/978-3-642-22362-4\\_18](https://doi.org/10.1007/978-3-642-22362-4_18)
- Mcdaniel, L., Jackson, A., & Gaudet, L. (2009). Can “withitness skills” be applied to teaching with laptops? In *American Journal of Business Education* (Vol. 2, Issue 4).

- 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. <https://doi.org/10.1007/s11412-006-9005-x>
- Molenaar, I., & Knoop-Van Campen, C. A. N. (2018). How teachers make dashboard information actionable. *IEEE Transactions on Learning Technologies*, 12(3), 347–355. <https://doi.org/10.1109/TLT.2018.2851585>
- Navarro, D. (2015). *Learning statistics with R: A tutorial for psychology students and other beginners (Version 0.6)* (Version 0.). null.
- Park, H., & Zhang, J. (2022). *Learning analytics for teacher noticing and scaffolding: Facilitating knowledge building progress in science learning analytics for teacher noticing and scaffolding: Facilitating knowledge building progress in science*. June.
- Ponce-López, V., Escalante, H. J., Escalera, S., & Baró, X. (2015). *Gesture and Action Recognition by Evolved Dynamic Subgestures*. 129.1–129.13. <https://doi.org/10.5244/c.29.129>
- Prieto, Kshitij, S., & Dillenbourg, P. (2015). Design for teaching and learning in a networked world. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9307, 613–616. <https://doi.org/10.1007/978-3-319-24258-3>
- Prieto, Sharma, K., Kidzinski, Ł., & Dillenbourg, P. (2018). Orchestration load indicators and patterns: In-the-wild studies using mobile eye-tracking. *IEEE Transactions on Learning Technologies*, 11(2), 216–229. <https://doi.org/10.1109/TLT.2017.2690687>
- Prieto, Villagrà-Sobrino, S., Jorrín-Abellán, I. M., Martínez-Monés, A., & Dimitriadis, Y. (2011). Recurrent routines: Analyzing and supporting orchestration in technology-enhanced primary classrooms. *Computers and Education*, 57(1), 1214–1227. <https://doi.org/10.1016/j.compedu.2011.01.001>
- Race, P. (2000). *500 tips on group learning*. Kogan Page.
- Rummel, N., Deiglmayr, A., Spada, H., Kahrimanis, G., & Avouris, N. (2011). Analyzing collaborative interactions across domains and settings: An adaptable rating scheme. In S. P. et al. (Ed.), *Analyzing Interactions in CSCL, Computer-Supported Collaborative Learning Series*. Springer Science+Business Media. [https://doi.org/10.1007/978-1-4419-7710-6\\_17](https://doi.org/10.1007/978-1-4419-7710-6_17)
- Snoeyink, R. (2010). Using video self-analysis to improve the “Withitness” of student teachers. *Journal of Digital Learning in Teacher Education*, 26(3), 101–110. <https://doi.org/10.1080/10402454.2010.10784641>
- Soller, A., Martínez-Monés, 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, 261–290.
- Spikol, D., Ruffaldi, E., Landolfi, L., & Cukurova, M. (2017). Estimation of success in collaborative learning based on multimodal learning analytics features. *Proceedings - IEEE 17th International Conference on Advanced Learning Technologies, ICALT 2017*, 269–273. <https://doi.org/10.1109/ICALT.2017.122>
- Stadler, S., Kain, K., Giuliani, M., Mirnig, N., Stollnberger, G., & Tscheligi, M. (2016). Augmented reality for industrial robot programmers: Workload analysis for task-based, augmented reality-supported robot control. *25th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN 2016*, 179–184. <https://doi.org/10.1109/ROMAN.2016.7745108>
- Stahl, G. (2005). Group cognition in computer-assisted collaborative learning. *Journal of Computer Assisted Learning*, 21, 79–90.
- Stanislaw, H. (1999). Calculation of signal detection theory measures. In *Behavior research methods, instruments, & computers* (Vol. 3).
- Stroch, N. (2001). *How collaborative is pair work? ESL tertiary students composing in pairs*.
- Thalemann, S., & Strube, G. (2004). Shared knowledge in collaborative problem solving: Acquisition and effects. *Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society, September*, 1333–1338.
- Tindale, R. S., Kameda, T., & Hinsz, V. B. (2003). Group decision making. In M. A. Hogg & J. Cooper (Eds.), *SAGE handbook of social psychology* (pp. 381–403). SAGE.
- van Leeuwen, A., & Rummel, N. (2020). *Comparing teachers' use of mirroring and advising dashboards*. 26–34. <https://doi.org/10.1145/3375462.3375471>
- van Leeuwen, A., & Janssen, J. (2019). A systematic review of teacher guidance during collaborative learning in primary and secondary education. *Educational Research Review*, 27(September), 71–89. <https://doi.org/10.1016/j.edurev.2019.02.001>
- van Leeuwen, A., & Rummel, N. (2019). Orchestration tools to support the teacher during student collaboration: A review. *Unterrichtswissenschaft*, 47(2), 143–158. <https://doi.org/10.1007/s42010-019-00052-9>

- van Leeuwen, A., Rummel, N., & van Gog, T. (2019). What information should CSCL teacher dashboards provide to help teachers interpret CSCL situations? *International Journal of Computer-Supported Collaborative Learning*, 14(3), 261–289. <https://doi.org/10.1007/s11412-019-09299-x>
- 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(6), 1499–1514. <https://doi.org/10.1007/s00779-013-0751-2>
- Vieira, C., Parsons, P., & Byrd, V. (2018). Visual learning analytics of educational data: A systematic literature review and research agenda. *Computers and Education*, 122, 119–135. <https://doi.org/10.1016/j.compedu.2018.03.018>
- Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers and Education*, 46(1), 71–95. <https://doi.org/10.1016/j.compedu.2005.04.003>
- Wise, A. F. (2018). Learning analytics: Using data-informed decision-making to improve teaching and learning. In *Contemporary technologies in education: Maximizing student engagement, motivation, and learning* (pp. 119–143). Springer International Publishing. [https://doi.org/10.1007/978-3-319-89680-9\\_7](https://doi.org/10.1007/978-3-319-89680-9_7)
- Wise, A. F., & Jung, Y. (2019). Teaching with analytics: Towards a situated model of instructional decision-making. *Journal of Learning Analytics*, 6(2), 53–69. <https://doi.org/10.18608/jla.2019.62.4>
- Wolff, C. Eliza., Boshuizen, H. P. A., & Jarodzka, H. M. (2016). Revisiting “withitness”: Differences in teachers’ representations, perceptions, and interpretations of classroom management.
- Wolff, C. E., Jarodzka, H., & Boshuizen, H. P. A. (2020). Classroom management scripts: A theoretical model contrasting expert and novice teachers’ knowledge and awareness of classroom events. *Educational Psychology Review*, 33(1), 131–148. <https://doi.org/10.1007/s10648-020-09542-0>
- Worsley, M., Anderson, K., Melo, N., & Jang, J. (2021). Designing analytics for collaboration literacy and student empowerment. *Journal of Learning Analytics*, 8(1), 30–48. <https://doi.org/10.18608/jla.2021.7242>

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