



Do Teachers Find Dashboards Trustworthy, Actionable and Useful? A Vignette Study Using a Logs and Audio Dashboard

Reet Kasepalu¹ · Pankaj Chejara² · Luis P. Prieto¹ · Tobias Ley¹

Accepted: 20 April 2021 / Published online: 27 April 2021
© The Author(s) 2021, corrected publication 2021

Abstract

Monitoring and guiding multiple groups of students in face-to-face collaborative work is a demanding task which could possibly be alleviated with the use of a technological assistant in the form of learning analytics. However, it is still unclear whether teachers would indeed trust, understand, and use such analytics in their classroom practice and how they would interact with such an assistant. The present research aimed to find out what the perception of in-service secondary school teachers is when provided with a dashboard based on audio and digital trace data when monitoring a collaborative learning activity. In a vignette study, we presented twenty-one in-service teachers with videos from an authentic collaborative activity, together with visualizations of simple collaboration analytics of those activities. The teachers perceived the dashboards as providers of useful information for their everyday work. In addition to assisting in monitoring collaboration, the involved teachers imagined using it for picking out students in need, getting information about the individual contribution of each collaborator, or even as a basis for assessment. Our results highlight the need for guiding dashboards as only providing new information to teachers did not compel them to intervene and additionally, a guiding dashboard could possibly help less experienced teachers with data-informed assessment.

Keywords Collaborative analytics · Teacher support · Teaching augmentation · Monitoring students · Actionability · Design-based research · Learning analytics

1 Introduction

Although collaborative learning (CL) shows many benefits, simply having students work in a group does not automatically result in students' deep learning and construction of knowledge (Stover & Holland, 2018). But even if teachers are in the same classroom as the students during a CL activity, they still have little to no information about what is happening

✉ Reet Kasepalu
reetkase@tlu.ee

¹ Department of Educational Sciences, Tallinn University, Tallinn, Estonia

² Department of Digital Technologies, Tallinn University, Tallinn, Estonia

in the individual groups as their attention is dispersed among several groups. It is known that one teacher can moderate about four groups of three to four students in a face-to-face CL setting. If we go above those numbers, teachers may not be able to monitor and take into consideration the full complexity of group and individual functioning at the same time (Schwarz & Asterhan, 2011). The average class size according to the OECD in 2017 ranged from 15 to 33, meaning that one teacher would need to monitor and guide four to eight groups of quartets on average in a CL scenario. We should not expect teachers to be able to monitor the progress of each student.

In our study, we are focusing on the perceptions of in-service teachers using a learning analytics (LA) technology in the form of a dashboard. A teacher would not be able to provide support to students when she/he is not fully aware of the situations and progress in the collaborative classroom. Today, LA technology often just analyses logs from systems, which provides only half of the picture in face-to-face collaboration, whereas multimodal LA uses multiple data sources (video, audio, sensors) to bring additional data about the physical interactions to stakeholders and teachers (Lang et al., 2017). We suggest a CL dashboard with two data sources (audio and log data) could be of help here. The paper at hand reports on a vignette study with 21 practising secondary school teachers and focuses on their perceptions of the dashboard, which has not been thoroughly researched. We want to know if the teachers trust the dashboards and therefore also possibly grow more confident in their decisions.

2 Related Work

2.1 Teachers' Decision Making in Class

To provide effective support, teachers have to be constantly aware of students' activities and make fast decisions about which group to monitor and visit to offer support (Kaendler et al., 2015). The teacher's knowledge of the situations occurring in the classroom in real time is *situational awareness* (Wolff et al., 2020), which could be increased with the help of learning analytics and thus increase the effectiveness of the decisions they subsequently make (i.e., who needs support) (Verbert et al., 2014). When studying relevance using the conceptual framework for data-driven *decision making*, the question is if the data is relevant for the decision-maker for his/her practice. (Gill et al., 2014). Take CL as an example: teachers need to monitor several groups of students at the same time, make decisions on the spot on who to support and who to let work independently and gather information for feedback and reflection after the collaboration. Research shows that there needs to be *trust* towards the presented measures in the analytics (Wise & Schwarz, 2017).

2.2 Enhancing Awareness Using Learning Analytics

Multimodal LA uses different sources of data (Lang, 2017) to inform a *dashboard* that typically has the goal of promoting awareness, reflection, and sense-making (Verbert et al., 2014). Teachers today are in need of a dashboard which can increase their sensemaking in real time (An et al., 2020). Different dashboards have been created as a response to this need: using multi-tabletops (Martinez Maldonado et al., 2012) or wearable displays for teachers (Holstein et al., 2018), ambient awareness tools (Alavi & Dillenbourg, 2012; An et al., 2018) or ambient awareness tools combined with audio (Bakker et al., 2013). We

are proposing that a LA dashboard could help the teacher during CL taking over some of the tasks (monitoring) and the teacher taking others (offering help), creating a hybrid (human-machine) intelligence system. As it is not clear whether and how such a hybrid system could work effectively, the effect on the *decision-making process* of the teachers should be studied. There is already initial evidence that mirroring teacher dashboards are perceived by teachers as helpful and insightful (van Leeuwen & Rummel, 2020). Chounta and Avouris (2016) focused on usability and *usefulness* in their study on CL.

2.3 The Actionability of Real-Time Teaching Augmentation

In addition to confirming the initial evidence of being useful, research suggests that it is important to study how teachers make use of the dashboard and whether the dashboard indeed increases their situational awareness of the situation (van Leeuwen & Rummel, 2020). Thus adding *actionability* (understood as data that initiates a curative action (Jørnø & Gynther, 2018)) and *novelty* as items under study. Especially less experienced teachers might find themselves feeling uncertain during a CL class (Wolff et al., 2020), but teachers feeling confident in their practice can help reduce teacher stress while improving student outcome (Herman et al., 2018), which is why we will be studying the teachers' level of *confidence*. To help teachers with timely interventions in class, we are talking about *real-time teaching augmentation*. This is defined as something that enhances the teacher's didactic competencies in action with the goal of extending the teacher's awareness during a lesson (An et al., 2020). Although different dashboards have been developed, the effect the dashboard has on the teacher's decision-making process and overall awareness of the collaboration process is unknown.

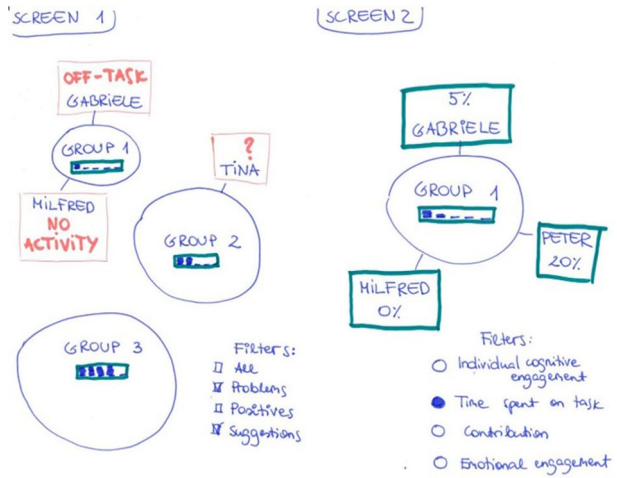
3 Design-Based Research

This study is part of a design-based research which has undergone three iterations. The feedback from end-users has been sought in each iteration to improve the final product, meaning that we seek to involve teachers into the design and research process from early on (Ley et al., 2009). The previous two iterations are briefly described below to provide the reader with an overview of the process and decisions made before introducing the principal focus of the study: iteration 3.

3.1 Iteration 1 Needs Analysis and Paper Prototype

The first iteration involved conducting a needs analysis and presenting eight in-service teachers with a paper prototype (see Fig. 1) to encourage teachers to discuss their concrete needs rather than abstract wishes (Ley et al., 2009). The interviewed teachers in the first iteration reported on wanting to know about the "power dynamics" of the group (who was more dominant, passive in the group); additionally, they were interested in knowing what the individual contribution of each student within the group is "*to have an objective basis for assessing the students*" (Kasepalu et al., 2019). Research suggests that social loafing (Chen et al., 2018), passivity (Koivuniemi et al., 2018) and dominance issues (Chen et al., 2018) are common problems teachers face in CL. This provided us with the idea of collecting data about the verbal and written contribution during a collaborative writing task,

Fig. 1 Paper prototype used in iteration 1



as evidence about the power dynamics as well as the quantified effort of each individual group member.

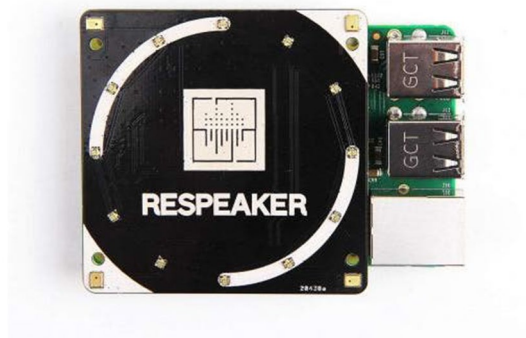
3.2 Iteration 2 CoTrack and Learning Task

3.2.1 CoTrack Dashboard

In the second iteration, we built a research prototype -CoTrack- (see Fig. 2) to capture data from face-to-face CL settings. Our prototype design was guided by the needs identified in the first iteration of the research in addition to research findings in the field which investigated the role of speaking behaviour in understanding collaboration. The three principles guiding the design process were: user-centeredness, basing it on theory and aiding the decision-making process (Liu & Nesbit, 2020).

Weinberger and Fischer (2006) identified the amount of participation as a key quantitative measure in CL. As the learning task in our research required students to discuss and

Fig. 2 Cotrack prototype



create a shared document, we decided to consider the students' participation in terms of their speaking and writing. This decision led us to use audio and logs data in our prototype, keeping the technological complexity at minimum compared to other data sources (e.g., video, eye-gaze). Our decision to use audio was also based on the research in the field that has found verbal participation (e.g., speaking time, the amount of words) as an indicator of collaborative behaviour (Reilly & Schneider, 2019). We used Etherpad (an open-source real-time collaborative editor) to allow students to write collaboratively.

We borrowed the idea of CoTrack from Noel et al. (2018) to capture participants' interaction in face-to-face settings, allowing us to obtain their amount of participation. CoTrack is a Raspberry Pi-based prototype with omnidirectional microphones. It detects the presence of sound using the Voice Activity Detection (VAD) algorithm and uses the Direction of Arrival (DoA) algorithm to estimate the direction of sound. The detected directions were then mapped to the corresponding students according to their sitting position around the prototype. This mapping provided us the speaking time and number of speaking turns for each student. From Etherpad logs, we extracted the number of characters added or deleted by each student in the group. In our prior work, we found that our prototype's simple features (e.g., speaking time, number of characters) were significant in estimating collaboration quality (further details, see Chejara et al., 2020). Figure 3 illustrates a part of CoTrack's dashboard showing speaking time and social network, which we presented to teachers to investigate their decision-making process during CL.

3.2.2 Learning Task

To test the prototype and our initial hypothesis, a case study was conducted in an upper secondary school's biology course in autumn 2019. The learning activity was co-designed by a teacher and the main author who has an education science background. The learning task for the students was to compile a code of ethics to be used for growing genetically

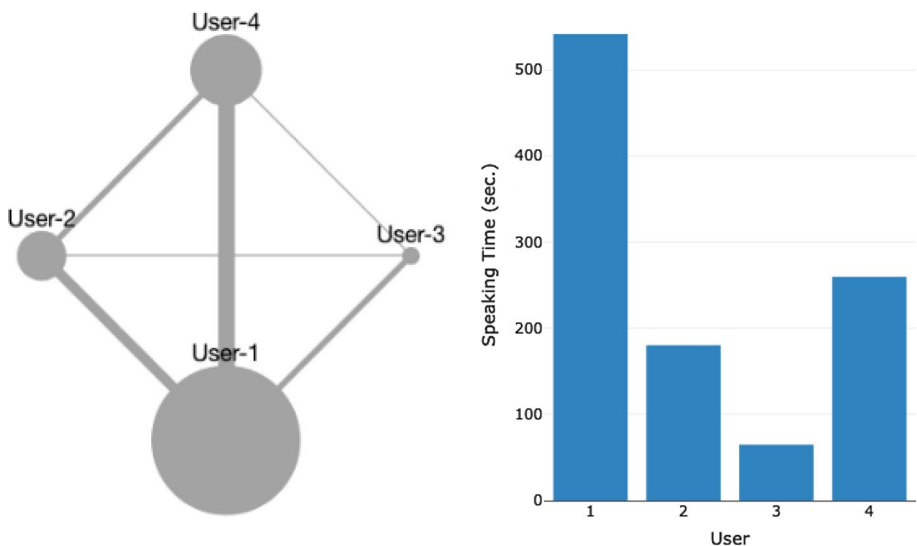


Fig. 3 Cotrack dashboards: social network analysis on the left and comparison of speaking times on the right

mutated crops in agriculture. The students had been discussing genetic mutation for four lessons and now had to apply their knowledge in a new context—writing a code of ethics. Their task was to compile at least 5 rules for the producers to follow when growing genetically mutated crops. Each group had to hand in one rulebook, which were presented and discussed frontally in the next lesson. 10 students took part in the CL activity while two researchers were present in the classroom for data collection purposes. Prior to the data collection, the aims of the study were introduced, the consent forms were filled by the students and the teacher. During the activity, audio- and log data together with video data for ground truth were collected (see Fig. 4 for the set-up).

The teacher in the case study said that the dashboard seemed useful for her everyday practice, but she would rather use it with bigger classes. The provided data seemed relevant to her, the spoken contribution offered confirmatory information for the teacher, whereas the individual written contribution of each student was new information. With the introduction of the dashboard, the teacher's confidence in assigning the dominant and passive members of the group increased. The teacher assessed the overall collaboration of the group used for the vignette design as effective. The question arose if the findings of our case study are generalizable to other teachers. In the quest to understand this and investigate further the relevance: trust, novelty, usefulness; actionability and confidence, we set up a third iteration—the study which we will focus on for the remainder of the paper.

4 Method

The goal of the study is to get an insight into the teachers' decision-making process and overall awareness of the collaboration process when presented with a dashboard with two different data sources.

The following research questions serve to investigate the decision-making process and level of awareness the teachers have of the students' collaboration using a dashboard:

- RQ1 Do the teachers gain confidence in their decisions about the power dynamics of the group after being provided with the information from the dashboard?

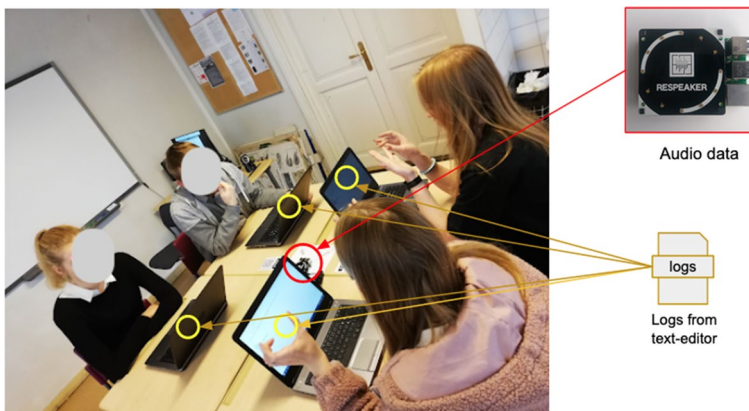


Fig. 4 Learning task

- RQ2 How relevant is the dashboard for the teachers?
 - RQ2.1 How much do the teachers trust the data provided in the dashboard?
 - RQ2.2 Do the teachers gain new information from the dashboard? What?
 - RQ2.3 Do the teachers see the information provided by the dashboard as useful in their practice? What for?
- RQ3 Is the information provided by the dashboard actionable? What type of interventions would the information prompt the teachers to take?
- RQ4 Do the participating teachers' perceptions differ according to the level of experience? How?

In RQ1, we wanted to know if providing teachers with a dashboard would help them gain confidence in their decision-making process when indicating which student is dominating or passive. RQ2 investigates the relevance of the dashboard [using the conceptual framework for data-driven decision making (Gill et al., 2014)] for teachers in terms of *trust, novelty, and usefulness*. For studying RQ2.1, the teachers were asked to rate how much they thought the dashboard reflected the reality of what was happening in the group work. The scale was on a continuum from zero (representing no connection to reality) to one hundred (representing reality like a mirror) and the teachers were asked to pinpoint how trustable the information provided by the prototype seemed. Intending to see whether the dashboard provides new information (RQ2.2), we asked that directly from all participating teachers. To find out whether the dashboard with two data sources would be useful (RQ2.3), we showed the participants the vignette, introduced the dashboard and let them decide whether they would see this as something that could be useful in their own classroom and if this was the case, then for what? Furthermore, we wanted to know what the teacher would decide to do if she had had the information provided in the dashboard during the lesson seen from the video (RQ3). Actionability was studied as a separate factor as it is highly important on the road towards being actively used in everyday practice (Gill et al., 2014). Furthermore, we divided the participating teachers into two experience groups and looked at whether differences in perceptions existed based on the Theoretical Model Contrasting Expert and Novice Teacher's Knowledge and Awareness of Classroom Events (Wolff et al., 2020).

4.1 Data Collection

4.1.1 Procedure

The 21 interviewed teachers watched an eight-minute compilation of an authentic CL scenario (originating from the case study in iteration 2). The method is similar to the vignette study carried out by Van Leeuwen et al. (2014) with the difference of showing only the collaboration of one group. Our vignette follows the protocol of replay enactments that is grounded in prior research and enables researchers to put together parts of recorded experiences to even create new trajectories which are still grounded in real experiences (Holstein et al., 2020). After seeing the video, the participating teachers were asked to answer three questions: (1) *was someone dominating? who?* (2) *was someone being passive? who?* (3) *was the group work effective in your opinion?* After each of these questions, the interviewee indicated how confident she was in her answer on the Likert scale from strongly confident = 5, confident = 4, neutral = 3, not confident = 2, to strongly unsure = 1 (Çatma &

Corlu, 2016). In the next phase, the interviewee was shown collaboration analytics in the form of a dashboard (the dashboard was introduced in iteration 2 in Fig. 3). The dashboard was based on the authentic CL activity from iteration 2. After that, the three questions were asked again with the indicating confidence level on a scale from 5 to 1. Having completed this, the teacher answered questions about the relevance (usefulness, novelty, trust) and actionability of the prototype (for the interview guide, go to Appendix 1).

4.1.2 Participants

The vignette study was conducted in face-to-face settings with twenty-one teachers, all teachers were interviewed in Estonian. Invitations to participate in the study were sent to twenty upper-secondary schools, only female teachers responded and were interested in taking part in the study. This is not surprising as there are more female teachers in Estonia than male teachers (86% female teachers in 2017). The vignette was shown to twenty-one female in-service upper-secondary school teachers. The teachers were informed that they were going to see an eight-minute video of a thirty-minute task, with the learning task described to the interviewed teachers before watching the video. In addition to this, it was possible for the participants to ask specifying questions before or after watching the video. The anonymity of the teachers is assured by referring to the authorship of the quotes in the results chapter with a number.

4.1.3 Instrument

A vignette study was chosen to investigate if the data provided in the dashboard is relevant to teachers. Research shows that when teachers have been presented with vignettes, the ability of the involved teachers to spot passive behaviour increased, they became more actionable with problematic students and were able to explain their actions (Van Leeuwen et al., 2014). To design the vignette, we interviewed the teacher in the case study described in iteration 2 just after the activities took place to get her expert opinion of the dynamics of the experienced CL task. We put together an eight-minute video compilation characteristic of the group dynamics according to the classroom teacher with one more dominant and one rather passive member (see Table 1 for the description of the situations).

The vignette intended to imitate an authentic classroom situation where the teacher is walking around the classroom and stopping by each group for some time, thus experiencing only glimpses of the overall collaboration. We opted to set up the study so that from the total 30 min of a collaborative task, the teacher was able to only observe a portion of the work done by each group (similar to Van Leeuwen et al., 2014). Thus, at the same time losing insight into the process of collaboration of the other groups (as it happens in an authentic classroom setting, as the teacher goes over to monitor and support different groups).

4.2 Data Analysis

For qualitative analysis in the first cycle in iteration 3, structural coding was implemented, which is especially suitable for studies with multiple participants following a semi-structured data-gathering interview. The interviews were transcribed with the help of an automated transcription tool available at <http://bark.phon.ioc.ee/webtrans/> (Alumäe & Tilk, 2018), then further corrected by the main author. The qualitative software NVIVO was used to manage and code the transcripts, with the interviews initially coded into three

Table 1 Description of the collaboration of the group shown in the video compilation

Situation 1	Situation 2	Situation 3
No. 3 trying to collaborate and participate, no. 4 trying to encourage no. 3	No. 1 and no. 4 discussing in a pair, no. 3 idle	No. 4 addressing no. 3 for advice, no. 3 does not engage

themes: novelty, usefulness, actionability. In the second cycle, thematic analysis was conducted. Thematic analysis of the theme usefulness resulted in answering RQ2.3, thematic analysis of the theme actionability answered RQ3. The teachers are labelled T1–T21 to illustrate the authorship of the quotes hereon.

For the analysis of quantitative data, RStudio was used to answer RQ1 (confidence), RQ2.1 (trust) and RQ2.2 (novelty). To study RQ1, a paired samples T-Test with the alpha level 0.05 was conducted between the level of the reported confidence before and after being introduced to the dashboards. To study the impact of new information (RQ2.2), a comparative analysis was carried out. Descriptive statistics were implemented, and a histogram compiled to visualise to which extent the teachers trust the data (RQ2.1). The results from the qualitative analysis from the themes usefulness and actionability were inserted as new variables (overview of contribution, basis for assessment etc.) to compare the proposed actions between teachers' other characteristics. As the level of agreement between teachers was rather low, we wanted to see whether the setup of the study could have been an issue for the teachers. We looked through the transcriptions to see if any of the teachers commented on not having enough data for decision making and did not find such comments; hence it seems that the length of the 8 min video for decision making purposes was not a problem for the teachers. To answer RQ4, we divided the teachers into two groups: less experienced teachers (up to 5 years of experience) and more experienced teachers (5+ years of experience) and carried out a contrasting analysis, going back to the interviews for examples.

5 Results

5.1 Do Teachers Gain Confidence in Their Decisions about Power Dynamics? (RQ1)

To test the effect on confidence, we compared the confidence rating the teachers had verbalised before and after inspecting the dashboard. The level of confidence when deciding about the power dynamics of the group increased with the introduction of the dashboard from 4.1 to 4.52 (dominance, see Fig. 5); from 4.19 to 4.52 (passivity) and from 4.0 to

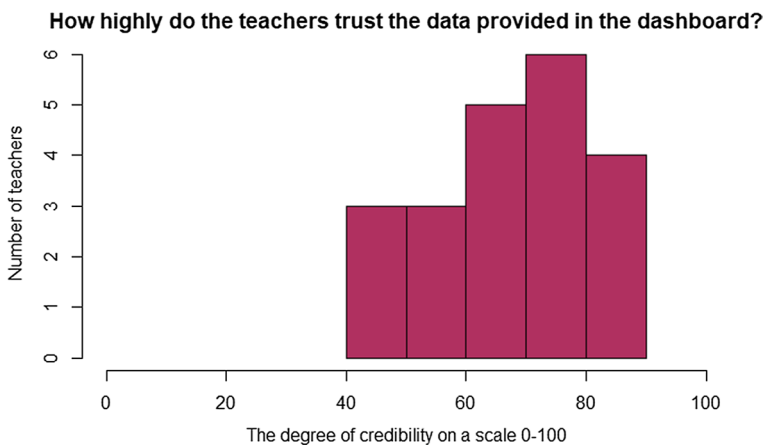


Fig. 5 Trustworthiness of the dashboard data in the teachers' opinion

4.29 (effectiveness). A paired samples T-Test (alpha level 0.05) confirmed a statistically significant difference in means for dominance $t(20) = 3.286$, $p = 0.004$, and for passivity $t(20) = 2.092$, $p = 0.049$. This means that the teachers collectively grew more confident in their assessment of assigning the most dominant and most passive person in the group after having been introduced to the dashboard. As the teachers grew more confident in their decisions about the power dynamics of the group after seeing the dashboard, this might be an indication that the dashboard could help the teachers make data-driven decisions.

As we had asked teachers whether the collaboration had been effective or not after looking at the video, the Fleiss Kappa for interrater reliability was computed resulting in a fair agreement (0.36). After looking at the dashboards, it resulted in a moderate agreement (0.49). Examining the level of agreement, it does stand out that although the teachers all saw the same video, their perception of the quality of collaboration diverged quite a lot, this meaning that the understanding of what it means to have an effective collaboration varies from teacher to teacher. Before seeing the dashboards, 14 out of 22 teachers evaluated the collaboration as not effective, and this was not significantly affected by the dashboard.

5.2 Is the Dashboard Relevant for Teachers? (RQ2)

5.2.1 RQ2.1 Trust

When introducing the dashboard to each participating teacher, they were informed that the ambient noise could have affected the outcome of the analysis. The participating teachers ($N = 21$) seemed to trust the information provided by technology, the mean on a scale from zero to one hundred is 69 with a range of 40–90 and a standard deviation of 15 (see Fig. 5), the distribution exhibits a positive skew. [T16] commented: “knowing how the machine operates, that the background noise is also computed in right now. Basing it on my experience, I would say it is still quite true, I feel confident in saying it is around eighty per cent”, which was the mode in this study. Teachers trusted the writing logs more than the audio data, but they did imply that having both data sources was valuable “if you have the speech and the writing on top of it, then it gives adequate feedback, which you would not have if only seeing the written contribution” [T19]. The teachers also asked to see data about the content of the talk making it crucially important to have an accurate audio data gathering device.

5.2.2 RQ2.2 Novelty

For some teachers, the dashboard offered new information ($N = 14$), other teachers ($N = 7$) said that the information confirmed the knowledge they had already obtained from the video, consolidating their decision made after the video. Five teachers reported on attaining new information from the log data saying that “I would not have thought that she had managed to write that much” [T5] or struggling to “know who had written the most” [T11]. Many teachers ($N = 12$) reported on getting new information from the audio data saying that “I had no idea number 4 had been so active” [T9], because “when I saw her, she did not speak, how can that be?” [T20]. This implies that teachers tend to draw conclusions based on limited data, which might stem from stereotyped decision patterns and/or not always be diagnostic for the overall collaboration. The teacher in the case study highlights the process that often occurs in CL “I usually make up my mind based on the last part of the collaboration, things I see and hear from the students, this is

also how I assess students, but I guess here I had not noticed how much he had actually tried to collaborate.” It seems this teacher focuses a lot of their attention on the product and the end of the process. Even the teacher who had been present in the classroom during the lesson had obtained new information from the dashboard. As the teacher from iteration 2 knew that the theoretical knowledge of one student was more limited, she did not notice how much the student had in fact been trying to cooperate with others.

5.2.3 RQ2.3 Usefulness

The teachers ($N=8$) indicated that using this type of prototype could have a stimulating effect on the students during the collaborative task. They hypothesized that “the students could be more motivated to actually start the group work just to see the graphs afterwards” [T8] and that it “would bring excitement into the classroom” [T10] alleviating the need for constant monitoring during the activity. Additionally, they ($N=8$) suggested using the collected information as the grounds for assigning group membership before the next collaborative activity “to put different students together who complement and could learn from each other” [T1], aiding the teacher in the planning phase. Another recurring answer ($N=10$) showed teachers using it as a basis for assessment where the dashboard “provided clarity for assessment” [T5] and “objective feedback stripped of emotions” [T13]. Some teachers ($N=5$) would prefer to use this as a solely formative assessment tool and to get an overview of the contribution of each student in the group, whereas others ($N=5$) would be ready to use it as basis for assessment. The teachers indicated that grading CL can be difficult for teachers as the teacher “can let the other group members assess one another, but in the end, it is one’s word against another, who to believe?” [T4]. The dashboard could thus possibly “create a sense of security” [T20] for the teachers. It was also expressed ($N=6$) that the prototype provided the involved teachers “with an overview of what was actually happening within the groups” [T9], “a visual of how the students are working” [T1]; hence, aiding with the competence of monitoring. In addition to this, the teachers ($N=2$) mentioned it having a security camera effect, meaning that by simply using a group analytics device, the collaboration efficiency might increase because the students are trying harder. In addition to this, it was brought out by T3 that “only getting data about who was speaking would not give me much, I would only know that someone was talking. But if I also have the data on the writing process, I already have an overview of the role division within the group”, which indicates the importance of having a dashboard with different sources. The teachers ($N=14$) did express a preference for having the audio and writing data in one visual, so they could grasp in one glance how the group was doing.

The tool was considered time-consuming by some teachers and they brought out that they “would not want to work that much. It is easier to just observe the students and make them do self-reports, less work” [T2]. The time and extra work factor need to be taken into consideration when designing the tool further, the data gathering should be made as easy as possible for the teacher and the dashboards should provide valuable information at a single glance.

When teachers were asked: “To sum up, would you say this tool would be useful in your everyday practice at school?”, most of the teachers found the dashboards useful (19 out of 21). T17 indicated that she would only use this tool in a classroom setting once the accuracy is 90%, not before. T2 said she would not use it in her school setting

because all her students are so used to group work and collaboration that she did not see this as a necessity. Two teachers and the teacher from the case study argued that this tool would be useful with bigger groups as it is then more difficult for the teacher to monitor how all the students in the groups are doing and participating.

5.3 How Actionable is the Information Provided by the Dashboard? (RQ3)

After having shown the vignette and the collaboration analytics dashboard to teachers, we asked them to imagine being the responsible teacher in the video, and to describe how they would react. Most teachers ($N=19$) would use the information in some way during or after the CL activity. Eleven teachers would intervene by going and talking to whether the passive student directly or by first talking to the whole group with the passive member. [T8] specified that "if the information came so clearly into my hands during the lesson, then I would definitely intervene". Three teachers would use this information to rearrange groups. Two teachers would not talk to the concerned student immediately, but instead use this information to go closer to the students and observe them, making sure they notice the teacher's vicinity. Two teachers specifically said they would not do anything. To sum up, out of 21 teachers, 19 would use the information from the dashboard in either the planning, monitoring, supporting, or reflecting phase. In this hypothetical scenario, the dashboard seemed actionable, as the teachers said they would adapt their behaviour during or after the activity.

As we also wanted to see how the teachers would use the information during the CL session, we counted that out of 21 teachers there were 7 teachers who would not intervene during the session. To see the connection between actionability and new information, we marked all the teachers who had verbalized a specific intervention and we then carried out a comparative analysis. It can be perceived in Fig. 6 that six out of seven teachers who did not state a specific intervention during the lesson, were teachers who had gained new information from the dashboard. This suggests that the information does not necessarily need to be new to be actionable for the teacher, probably some other factor is of more crucial importance. Even more, it might be that surprising information

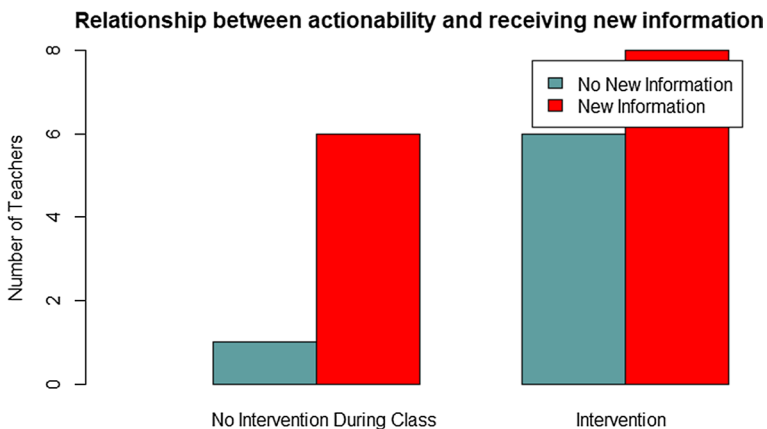


Fig. 6 Most of the non-intervening teachers were the ones having received new information

does not initiate an intervention because the teachers are not trained nor guided to use the information to intervene.

5.4 Do the Participating Teachers' Perceptions Differ According to the Level of Experience? How? (RQ4)

Based the Theoretical Model Contrasting Expert and Novice Teacher's Knowledge and Awareness of Classroom Events (Wolff et al., 2020), we wanted to check whether there were any noticeable differences in perceiving the collaboration that would be connected to the level of experience. We carried out a comparative analysis and there were three variables which stood out. Table 2 shows that less experienced teachers were more likely to assess the seen collaboration as effective, meaning that more experienced teachers might have higher expectations of the students' collaboration. More experienced teachers perceived the seen collaboration [T10] "as not effective", nor [T13] "intensive". However, a less experienced teacher T6 said for instance that "she would have been happy if her students had been discussing like this". Only one experienced teacher was ready to use the dashboard as a basis for grading, 10 were not. 4 out of 5 teachers ready to use the dashboard for grading were less experienced teachers (see Table 2). Less experienced teachers said that using the dashboard for grading [T4] "has potential, I do like the idea, especially if someone has fallen ill by the time of the presentation, then this provides me with some objective truth", [T5] "some grounds for grading that I could provide as evidence for the students". It can be inferred that less experienced teachers are having more trouble assessing collaboration and would potentially need some help in this area.

The third prominent variable emerging from the comparative analysis was readiness to change their minds. More experienced teachers were more ready to change their mind about the power dynamics of the group based on data presented to them. [T18] with almost 40 years of active teaching experience commented after seeing the dashboard: "Now I will have to say that student nr. 4 seems to have been more dominant. This graph shows me this. If I had my doubts beforehand, then now I am quite certain.". Most of the less experienced teachers were not affected by the data presented in the dashboard, e.g. [T7], "wow, here it shows that nr 4 had been talking a lot more, but I saw that she was totally silent in between "and thus opting to not change her mind about the most dominant student. It seems that experience helps to understand that a teacher's situational awareness is limited, which is where technology might help, for less experienced teachers this might need more persuasion.

Table 2 More experienced teachers are not ready to use the dashboard for grading, seem to have higher expectations of collaboration but will change their initial decision when presented with data

	Not effective/effective		Not use/use for grading		Change/not change initial decision	
Less experienced	4	6	6	4	1	9
More experienced	8	3	10	1	7	4

6 Discussion

The results from RQ1 showed that the involved teachers were rather confident in their initial decisions having only limited data. This might influence the perceived necessity of using LA. In our study, we focused on studying relevance from the viewpoint of the decision-maker; Gill et al. (2014) proposed diagnosticity as another important factor when looking into data-driven decision making. Would it possibly help the students in CL settings if teachers were to understand how they are currently making decisions based on nondiagnostic data every day (e.g., assessing the collaboration based on the end of the activity)? *Our implication for further studies* is testing the hypothesis that using technological assistants in the form of collaboration analytics (with diagnostic data) helps teachers monitor and support students much more efficiently and effectively.

As already mentioned, RQ2 focused on studying the relevance of the dashboard. Our study confirmed that the use of data by teachers for making decisions is useful. Nevertheless, some teachers also indicated that it is easier for them to solicit self-reported feedback from the group to find out what the individual contribution of each member was. As showing a dashboard possibly increases the orchestration load, which is the level of exertion a teacher dedicates to coordinating multiple learning processes (Prieto et al., 2018), the *general design principle* to follow is for the data to be demonstrated in a rather rustic and straightforward way. The orchestration load is higher with a bigger group of students, hence teachers might be more in need of this type of support with bigger classes. Although asking for self-report might be less time-consuming for some of the involved teachers, it has been shown that when comparing different feedback options, self-reported feedback was the only option to actually increase social loafing in a group (Suleiman & Watson, 2008). *Implication for further study* would be to compare the accuracy of students' self-report measures with the data derived from the dashboard, teacher assessment, peer assessment and video analysis or see if both kinds can be combined into a more reliable assessment. *The implication for pedagogical practice* is to combine another form of evaluation with self-assessment when assessing CL. When discussing *methodological implications*, the log and audio data will be combined into one dashboard to still offer multiple channels but at one glance. As the different sources of data were of interest to the teachers, others (e.g., eye gaze, facial expression recognition) could possibly be integrated into the dashboard. Additionally, a method of analysing the quality of collaboration will be incorporated into the dashboard to potentially give an objective basis for assessing collaboration. For a *general design principle*, the data gathering should be made as easy as possible for the teacher and the dashboards should provide valuable information at a single glance.

The notion of orchestration load could also be an important factor when looking at the novelty aspect. If the information given to the teachers is "too" new, it increases the orchestration load even more. As the results of our study revealed that gaining new information from the dashboard might decrease actionability, the system should be ready to provide some hints for intervention during the collaboration. Most of the tools offered today usually solely mirror information to teachers, however, the teachers are still left to decide how to interpret and act upon it (van Leeuwen & Rummel, 2019). *Implications for the design of collaborative learning dashboards.* As it seemed that new information might deter action in class, our plan is to implement different design principles for teachers with different experience to better support teachers in a CL classroom. When experienced teachers also benefit from a mirroring dashboard, less experienced

teachers seem to be needing a guiding system. As less experienced teachers were ready to use the dashboard for assessment purposes, the dashboard needs to be designed using a pedagogical model or theory.

Our study also exposed some differences in perceptions between less and more experienced teachers. Less experienced teachers may not have a clear set of criteria to assess CL by, they would consider using a dashboard to help them with assessment, but they seem to be less keen to change their minds about the power dynamics based on the seen data. The teaching augmentation framework suggests that teachers might need more scaffolding from the dashboard for the data to be actionable as too low-level data does not instigate an intervention (An et al., 2020). As an *implication for methodology*, a further experiment comparing teachers' behaviour in a control (no analytics used) and a test group (using guiding collaboration analytics) introducing experience as a variable under study could shed further light on this issue. This experiment could also consolidate or repudiate whether the dashboard is actionable in authentic settings.

The limitations of the study entail a small voluntary sample consisting of solely female teachers and basing the vignette study on opinions and hypothetical scenarios rather than authentic classroom situations. It needs to be considered that the sample for comparative analysis was ad hoc. Additionally, the teachers were only asked to observe one group at a time, more authentic results would have been yielded from a study using multiple groups in the vignette design. Our next steps in the design-based research include the development and validation of a model for assessing CL, as well as that of a flexible dashboard to offer strategies for teacher intervention.

7 Conclusion

This study presented how teachers perceive the relevance of a LA dashboard with two data sources in an authentic classroom setting. The involved teachers were rather confident in the information provided and hypothesized that it would prompt them to intervene during the collaboration in a variety of ways, for instance, teachers would use it to help solicit students in need of support, it could be used as a basis for assessment or even grading, it could also be used in further planning of CL groups and activities etc. It seems that the dashboard appeared relevant for the participating teachers and measured something meaningful to them.

The results from our study show that although new information enhances the situational awareness of the teachers, obtaining new information from the dashboard also seems to decrease actionability. We hypothesize that this possibly leaves the teacher, who is not able to notice the problem her/himself, in a situation where she/he still does not know how to react. Even after the dashboard has identified a specific issue, the teacher is still expected to carry out an intervention her/himself, without any support. This means that a mirroring dashboard might not be sufficient for the teachers' needs. Offering teachers a guiding LA dashboard, thus creating a hybrid human-machine intelligence, could help teachers in developing their CL management skills.

Appendix: Interview Guide

Intro

- (a) “The aim of the present research was to find out whether providing an analysis based on audio and digital trace data seems valuable for in-service secondary school teachers. First I am going to show you a video compilation of an authentic group work, the group work lasted 25min, you will see 8 min of it imitating a real classroom situation where it is not possible for the teacher to focus on one group for all of the lesson. When you watch the video, I want you to think about three things: who is dominating, if at all, who is being passive, if at all and to come to a conclusion whether this collaboration seemed effective to you or not. Do you have any questions?”
- (b) When you refer to the students, please refer to them as user 1, 3 and 4, here is a visual aid to help you talk about the students.

Video show the teacher a compilation of the discussion in group 2. Let the teacher ask questions when necessary.

Pre-test Let the teacher answer the following questions:

- (1) Do you think someone is dominating in the group? why? how sure are you? RQ1
- (2) Do you think someone is being inactive in the group? why? how sure are you? RQ1
- (3) Do you think the group is working effectively? how do you know? how sure are you? RQ2

Dashboard show the multimodal dashboards for group 2. In the four dashboards above you can see the analysis based on audio data, the two dashboards below are based on writing logs. Would you like to ask anything?

Post-test Let the teacher answer the following questions:

- (1) Do you think someone is dominating in the group? why? how sure are you? RQ1
- (2) Do you think someone who is being inactive in the group? why? how sure are you? RQ1
- (3) Do you think is the group is working effectively? how do you know? how sure are you? RQ2
- (4) Did you get any new information from the graphs? what?
- (5) Would you take any action or change your practice, on the basis of this information? what actions? during the process, after
- (6) How much do you think this graph reflects the reality? 0–100% why?
- (7) How helpful is the information for you? would you use it in your practice? ...

Thank the teacher.

Declarations

Conflict of interest There are no conflicts of interest to declare, this research has been partially funded by the European Union via the European Regional Development Fund and in the context of CEITER and Next-Lab (Horizon 2020 Research and Innovation Programme, Grant Agreements Nos. 669074 and 731685).

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

- Alavi, H. S., & Dillenbourg, P. (2012). An ambient awareness tool for supporting supervised collaborative problem solving. *IEEE Transactions on Learning Technologies*, 5(3), 264–274. <https://doi.org/10.1109/TLT.2012.7>
- Alumäe, T., & Tilk, O. A. (2018). Advanced Rich Transcription System for Estonian Speech. *Baltic HLT*.
- An, P., Bakker, S., Ordanovski, S., Taconis, R., & Eggen, B. (2018). ClassBeacons: Designing distributed visualization of teachers' physical proximity in the classroom. In *TEI 2018—Proceedings of the 12th international conference on tangible, embedded, and embodied interaction, 2018-Janua*, pp. 357–367. <https://doi.org/10.1145/3173225.3173243>
- An, P., Holstein, K., D'Anjou, B., Eggen, B., & Bakker, S. (2020). The TA framework: Designing real-time teaching augmentation for K-12 classrooms. arXiv.
- Bakker, S., Van Den Hoven, E., & Eggen, E. (2013). FireFlies: Physical peripheral interaction design for the everyday routine of primary school teachers. In *TEI 2013—Proceedings of the 7th international conference on tangible, embedded and embodied interaction*, 57–64. <https://doi.org/10.1145/2460625.2460634>
- Çatma, Z., & Corlu, M. S. (2016). How special are teachers of specialized schools? Assessing self-confidence levels in the technology domain. *Eurasia Journal of Mathematics, Science and Technology Education*, 12(3), 583–592. <https://doi.org/10.12973/eurasia.2016.1245a>
- Chejara, P., Prieto, L. P., & Ruiz-Calleja, A., Rodríguez-Triana, M.J., Shankar, S.K., Kasepalu, R. (2020). Quantifying Collaboration Quality in Face-to-Face Classroom Settings Using MMLA. In *26th international conference on collaboration technologies and social computing*. CollabTech 2020.
- Chen, J., Wang, M., Kirschner, P. A., & Tsai, C.-C. (2018). The role of collaboration, computer use, learning environments, and supporting strategies in CSCL: A meta-analysis. *Review of Educational Research*, 88(6), 799–884. <https://doi.org/10.3102/0034654318791584>
- Chounta, I. A., & Avouris, N. (2016). Towards the real-time evaluation of collaborative activities: Integration of an automatic rater of collaboration quality in the classroom from the teacher's perspective. *Education and Information Technologies*, 21(4), 815–835. <https://doi.org/10.1007/s10639-014-9355-3>
- Gill, B., Borden, B., & Hallgren, K. (2014). A conceptual framework for data-driven decision making. *Mathematica Policy Research*, 609, 1–13
- Herman, K. C., Hickmon-Rosa, J., & Reinke, W. M. (2018). Empirically derived profiles of teacher stress, burnout, self-efficacy, and coping and associated student outcomes. *Journal of Positive Behavior Interventions*, 20(2), 90–100. <https://doi.org/10.1177/1098300717732066>
- Holstein, K., Harpstead, E., Gulotta, R., & Forlizzi, J. (2020). Replay enactments: Exploring possible futures through historical data. In *DIS 2020—Proceedings of the 2020 ACM designing interactive systems conference*, pp. 1607–1618. <https://doi.org/10.1145/3357236.3395427>
- Holstein, K., Hong, G., Tegene, M., McLaren, B. M., & Alevin, V. (2018). The classroom as a dashboard. In *Proceedings of the 8th international conference on learning analytics and knowledge—LAK '18*, (c), 79–88. <https://doi.org/10.1145/3170358.3170377>
- Jørnø, R. L., & Gynther, K. (2018). What constitutes an 'Actionable Insight' in learning analytics? *Journal of Learning Analytics*, 5(3), 198–221. <https://doi.org/10.18608/jla.2018.53.13>
- Kaendler, C., Wiedmann, M., Rummel, N., & Spada, H. (2015). Teacher competencies for the implementation of collaborative learning in the classroom: A framework and research review. *Educational Psychology Review*, 27(3), 505–536. <https://doi.org/10.1007/s10648-014-9288-9>
- Kasepalu, R., Prieto, P. L., & Ley, T. (2019). Providing teachers with individual and group-level collaboration analytics : A paper prototype. In *International workshop on collaboration analytics: Making learning visible in collaborative settings*. Lyon: CSCL 2019. <https://collaborationanalytics.files.wordpress.com/2019/06/submission-2-reet.pdf>
- Koivuniemi, M., Järvenoja, H., & Järvelä, S. (2018). Teacher education students' strategic activities in challenging collaborative learning situations. *Learning, Culture and Social Interaction*, 19, 109–123. <https://doi.org/10.1016/j.lcsi.2018.05.002>

- Lang, C., Siemens, G., Wise, A., & Gašević, D. (2017). *Handbook of learning analytics. Handbook of learning analytics*. <https://doi.org/10.18608/hla17>
- Ley, T., Kump, B., Maas, A., Maiden, N., & Albert, D. (2009). Evaluating the adaptation of a learning system before the prototype is ready: A paper-based lab study (Vol. 5535, pp. 331–336). https://doi.org/10.1007/978-3-642-02247-0_32
- Liu, A. L., & Nesbit, J. C. (2020). *Dashboards for computer-supported collaborative learning. Advances in learning analytics*. Springer International Publishing. https://doi.org/10.1007/978-3-319-47194-5_5
- Martinez Maldonado, R., Kay, J., Yacef, K., & Schwendimann, B. (2012). An interactive teacher's dashboard for monitoring groups in a multi-tabletop learning environment. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7315 LNCS, 482–492. https://doi.org/10.1007/978-3-642-30950-2_62
- Noel, R., Riquelme, E., Lean, R. M., Merino, E., Cechinel, C., Barcelos, T. S., et al. (2018). Exploring collaborative writing of user stories with multimodal learning analytics: A case study on a software engineering course. *IEEE Access*, 6, 67783–67798
- Prieto, L. P., 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>
- Reilly, J. M., & Schneider, B. (2019). Predicting the quality of collaborative problem solving through linguistic analysis of discourse. In *EDM 2019 - Proceedings of the 12th international conference on educational data mining*, (Edm), 149–157.
- Schwarz, B. B., & Asterhan, C. S. (2011). E-moderation of synchronous discussions in educational settings: A nascent practice. *Journal of the Learning Sciences*, 20(3), 395–442. <https://doi.org/10.1080/10508406.2011.553257>
- Stover, S., & Holland, C. (2018). Student resistance to collaborative learning. *International Journal for the Scholarship of Teaching and Learning*, 12(2), 4. <https://doi.org/10.20429/ijstol.2018.120208>
- Suleiman, J., & Watson, R. T. (2008). Social loafing in technology-supported teams. *Computer Supported Cooperative Work*, 17(4), 291–309. <https://doi.org/10.1007/s10606-008-9075-6>
- Van Leeuwen, A., Janssen, J., Erkens, G., & Brekelmans, M. (2014). Supporting teachers in guiding collaborating students: Effects of learning analytics in CSCL. *Computers and Education*, 79, 28–39. <https://doi.org/10.1016/j.compedu.2014.07.007>
- 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. (2020). Comparing teachers' use of mirroring and advising dashboards (pp. 26–34). ACM, New York, NY, USA: *Proceedings of the 10th International Conference on Learning Analytics & Knowledge (LAK'20)*. <https://doi.org/10.1145/3375462.3375471>
- 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>
- 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., & Schwarz, B. B. (2017). Visions of CSCL: eight provocations for the future of the field. *International Journal of Computer-Supported Collaborative Learning*, 12(4), 423–467. <https://doi.org/10.1007/s11412-017-9267-5>
- 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>