Comparing Teachers' Use of Mirroring and Advising Dashboards

Anouschka van Leeuwen
Department of Education
Utrecht University
Utrecht, the Netherlands
a.vanleeuwen@uu.nl

Nikol Rummel
Institute of Educational Research
Ruhr-Universität Bochum
Bochum, Germany
nikol.rummel@rub.de

ABSTRACT

Teachers play an essential role during collaborative learning. To provide effective support, teachers have to be constantly aware of students' activities and make fast decisions about which group to offer support, without disrupting students' collaborative process. Teacher dashboards are visual displays that provide analytics about learners to help teachers increase their awareness of the situation. However, if teachers are not able to efficiently and effectively distill information from the dashboard, the dashboard can become an obstacle instead of an aid. In the present study, we compared dashboards that provide information (mirroring) to dashboards that provide information and alert the teacher to groups that are in need of support (advising). Teachers were shown standardized, fictitious collaborative situations on one of the types of dashboards and were asked to detect the group that was in need of support. The results showed that teachers in the advising condition more often detected the problematic group, needed less effort to do so, and were more confident of their decisions. The teacher-dashboard interaction patterns showed that teachers in the advising condition generally started by checking the given alert, but also that they tried to look at as much information about other groups as they could. In the mirroring condition, teachers generally started by examining information from class overviews, but did not always have time to check information for individual groups. These findings are discussed in light of the role of a teacher dashboard in teachers' decision making in the context of student collaboration.

CCS CONCEPTS

• Computer supported cooperative work • Visualization systems and tools • Empirical studies in HCI

KEYWORDS

cooperative/collaborative learning, elementary education, human-computer interface, improving classroom teaching, teaching/learning strategies

ACM Reference format:

Anouschka van Leeuwen, Nikol Rummel. 2020. Comparing Teachers' Use of Mirroring and Advising Dashboards. In *Proceedings of the 10th International Conference on Learning Analytics & Knowledge (LAK'20). ACM, New York, NY, USA, X pages.*
https://doi.org/10.1145/3375462.3375471

1 Introduction

Collaborative learning has been shown to be an effective method for learning [13]. Teachers play an essential role during collaborative learning [34]: they monitor the groups of students, and provide support concerning the task and the collaborative process when needed. To provide effective support, teachers have to be constantly aware of students' activities and make fast decisions about which group to observe and which group to visit to offer support [10] without disrupting students' collaborative process [15]. Recently, the idea of using teacher orchestration tools to aid teachers has gained considerable attention. The underlying idea is that orchestration tools making use of learning analytics techniques, such as dashboards that provide information about learners, can help teachers to increase their awareness of the situation and thereby increase the effectiveness of decisions they subsequently make concerning which group to support and what type of support to offer [26]. However, dashboards add an additional source of information in the already dynamic classroom. If teachers are not able to efficiently and effectively distill information from the dashboard, the dashboard can become an obstacle instead of an aid [8]. It is therefore highly important to also study how teachers make use of the dashboard and whether they indeed increase their awareness of the situation. In the present study, we examine how teachers interact with two types of dashboards, namely mirroring and advising dashboards, and investigate whether these teacher-dashboard interaction patterns can be related to the accuracy of teachers' detection of which collaborating groups are in need of support.

1.1 Teacher decision making and the role of teacher dashboards

Collaborative learning is the shared activity of two or more students towards a shared goal, and has been shown to lead to increased student learning [13]. When a classroom of students collaborates in small groups, there is a considerable amount of activity to monitor for the teacher, as both cognitive and social aspects of the collaborative process require the teacher's support [6, 27]. Maintaining overview during students' collaborate activity is a challenge for teachers [5], but a necessary skill for providing effective support [10]. By monitoring the collaborative activity and detecting groups that may need support, teachers can subsequently examine whether support is indeed necessary, and if so, select the appropriate pedagogical intervention [34]. Given the dynamic nature of the collaborative classroom, teachers are generally under pressure to keep up to date with all activity and have to continuously decide which group receives their attention at any given moment [6]. The teacher's decision about which group to interact with in the classroom has consequences for whether students display on-task or off-task behavior, not only for the group the teacher directly interacts with but also for the groups that are in close proximity to the teacher [2, 7]. It is therefore highly important that teachers detect in which groups they may be needed without disrupting students' collaborative processes [15]. It is this initial phase of detecting groups that may need support that we focus on in this study.

In response to the teacher's demanding task of monitoring and supporting student collaboration, modern technologies making use of learning analytics techniques have been developed to aid teachers [32]. The underlying idea of these technologies is that by offering information about learners, they can help teachers to increase their awareness of the activities in the classroom and the effectiveness of decisions they subsequently make. When collaboration between students is facilitated by computer software, digital traces of the students' activity are often automatically collected. The resulting data can be collected, analyzed, and displayed to inform the teacher. This process is an example of a larger body of work called learning analytics, which aims to collect information about learners with the goal of improving learning or the environment in which it occurs [14]. In this case, data is collected about learners to inform the teacher, who in turn can more effectively support students.

In the present study, we focus on teacher dashboards, by which we mean visual displays that provide teachers with information about their collaborating students [29, 32]. As described above, teachers must monitor students' activity, detect groups that may need support, and decide on subsequent action. Similarly, teacher dashboards can fulfill different functions that tie in to one of these phases of teacher decision making [25]. Mirroring dashboards provide information about learners to support monitoring of collaborative activity, but leave all subsequent detection and interpretation of relevant information to the teacher. Advising dashboards, on the other hand, provide alerts about groups that may be in need of support as well as advice about what problem the group could be facing (for an overview see [29]). Within these functionalities, dashboards may provide different types of information, such as the distinction between exploratory versus explanatory dashboards in [4]. To be clear, in this paper we focus on the function that the dashboard fulfills in supporting the teacher to interpret the information.

There is initial evidence that mirroring teacher dashboards are perceived by teachers as helpful and insightful (e.g., [12]), indeed offering teachers the information that allows them to look into the activity of all collaborating groups [3]. On the other hand, experimental studies that examine whether teachers' diagnosis of a situation improves do not always show significant effects (e.g., [28]). These findings thus point to the possible advantage of dashboards that fulfill more advanced roles such as advising dashboards, because they support teachers not only by providing information but also by supporting the detection of relevant events and the interpretation of those events. Unfortunately, research concerning advising dashboards is scarce [24]. The first aim of the present study is therefore to check whether advising dashboards indeed lead to more accurate detection of relevant events, which we will do by comparing teachers' detection accuracy for mirroring and advising dashboards.

1.2 Teacher-dashboard interaction patterns

Qualitative data concerning mirroring dashboards show that a possible explanation for the finding that detection accuracy does not always improve is that it is not self-evident that teachers are able to use mirroring dashboards to their advantage, for example because the amount of information is overwhelming [18, 33] or because shown information is interpreted differently than intended [28]. For both mirroring and advising dashboards, it can be said that these technologies add an additional source of information in the already dynamic collaborative classroom. How teachers make use of the dashboard directly precedes their interpretation of the situation and thus their subsequent decision making [35]. It is therefore highly important to study how teachers make use of the dashboard in terms of what information they look at and in which order they do so.

The study of how teachers interact with a dashboard is a form of learning analytics at a meta-level, because data is collected from teachers about how they interact with data that is collected about learners [31]. Although there are studies looking at dashboards with different functionalities, studies that have specifically looked into these teacher-dashboard interaction patterns mostly concern mirroring dashboards. These studies describe that teachers generally show the following behaviors to make sense of the information on the dashboard: they monitor the class and individual collaborating groups, for example by switching between class and group overviews on the dashboard; when they detect (potentially) relevant events they investigate more in detail to obtain a more specific diagnosis, for example by zooming in on specific information; and they proceed to direct communication with a group of students or an individual student to provide support if needed (e.g., [23, 26, 33]). These studies also showed that during the monitoring phase, teachers experienced time pressure and were not able to thoroughly process all information provided to them. These findings confirm the earlier mentioned mixed findings concerning quantified measures of teachers' detection of groups that might be in need of support, and strengthen the hypothesis that advising dashboards could be an aid to teachers. In particular, advising dashboards could make the monitoring phase more efficient by pointing to groups that show some form of deviating behavior, freeing time for teachers to process information about selected groups on the dashboard. To the best of our knowledge, only one other study explicitly compared dashboards with different functions, namely a mirroring and alerting dashboard [16]. In this study, it was found that only in the alerting condition, the teachers' feedback significantly influenced students' achievement. It could mean that teachers were better able to detect relevant information, upon which they could successfully act. There is thus a scarcity of studies that compare teacher-dashboard interaction patterns for multiple types of dashboards. Furthermore, there are no studies that we are aware of that draw on the relation between how a dashboard is used and whether specific interaction patterns are associated with the accuracy of the diagnosis of the situation.

1.3 The present study

To summarize, it is essential that teachers are up to date with the activities of collaborating students in order to stimulate effective collaboration. There is evidence that mirroring teacher dashboards can help teachers to do so, but experimental findings regarding effects on the accuracy of detecting groups in need of support are mixed. It is assumed that advising dashboards that provide more support in navigating information could lead to improved detection accuracy. More study is needed to test this hypothesis and to examine how teachers interact with mirroring and advising dashboards.

In the present study, we aim to extend existing research by comparing detection accuracy for mirroring and advising dashboards and by examining how teachers interact with these dashboards. Thus, in this study we combine measures of detection accuracy with an investigation into teacher-dashboard interaction patterns. The study has a controlled, experimental design in which teachers are provided with one of two types of dashboards that display information about collaborating students in a fictitious class. After interacting with a dashboard in each situation, teachers are asked to identify the collaborating group that is in need of support.

The following research questions were formulated:

1. What is the influence of mirroring and advising teacher dashboards on teachers' detection of collaborative groups that may need support?
2. What patterns of teacher-dashboard interaction occur for mirroring and advising dashboards?
3. Do teacher-dashboard interaction patterns differ for instances of accurate and non-accurate detection of groups that may need support?

2 Method

2.1 Design and Participants

An experimental study with a between-subjects design with two conditions was performed. The two conditions differed in the type of dashboard that was provided to the teachers, namely mirroring or advising dashboards. We investigated whether dashboard type influenced the accuracy of detecting information on the dashboard, and how teachers interacted with the dashboards in the two conditions.

The sample consisted of 35 participants, who were either pre-service primary school teachers or primary school teachers who had recently finished their teacher education. Participants signed up for the experiment voluntarily and received a monetary compensation for their participation. Participants were randomly distributed over the two conditions, leading to 17 in the mirroring condition (1 male) and 18 in the advising condition (2 male). Their mean age was 21.4 (SD = 2.2), and on average they had 30.4 months teaching experience in primary education (SD = 15.3). No significant differences between the conditions were found regarding these variables ($p > .6$ in both cases).

2.2 Materials and Procedure

The participants individually took part in the experiment, which was conducted fully computer-based. At the start of the experiment, participants watched a video that explained the procedure of the data collection and the layout of the dashboard they were about to interact with. The experiment continued with a number of questionnaires concerning teachers' background variables (age, sex, teaching experience) and teachers' experience with technology.

Participants were explained that their task was to imagine they were a teacher in a 4th grade class, in which dyads of students were collaborating on fraction assignments concerning the skills of naming fractions, simplifying fractions, and adding and subtracting fractions. The eight situations were derived from the existing software MathTutor [17], which is a program designed to practice mathematics (including fractions) for both individual and collaborative settings [20]. With MathTutor, both students have their own computer screen, but the interface they control is the same for the two members of the dyad. Each action on the interface is visible to the other dyad member, and by being seated next to each other and talking out loud, the dyad can discuss the assignments. Because MathTutor logs all student activity, these log-files can be processed and used as input for teacher dashboards.

It was explained to the participants that they would be shown eight dashboards that showed information about 5 dyads of students that collaborated through MathTutor. In the eight situations, which were shown in random order, the participants had to imagine that they consulted the dashboard to check whether any group might be in need of teacher support. It was stressed in the plenary explanation that participants should try to imagine they were in an actual classroom, even though the situations were fictitious.

The dashboard was designed based on a literature review and a co-design phase with teachers, see [30]. Information was available for the teachers to browse concerning the following six indicators: 1) the number of completed assignments, 2) the number of attempts a dyad needed to solve an assignment, 3) the chance that a dyad displayed trial-and-error behavior on an assignment, 4) the amount of talk for each dyad member, 5) dyads' proficiency on fraction skills, and 6) a display of a dyads' activity over time (explanation of these indicators follows below). The dashboards displayed information about the collaborating students at class and at group level. Figure 2 shows example dashboards for the mirroring and advising condition. On the left, buttons are displayed with the five dyad numbers. On the top row, six buttons are available for each of the six indicators. When a group button is clicked on, a group overview opens that displays information on all six indicators for that particular group (see Figure 2). When an indicator button is clicked on, a class overview opens that displays information concerning that indicator for each of the five groups. Figure 1 shows a more detailed image of the six indicators on a group overview page.

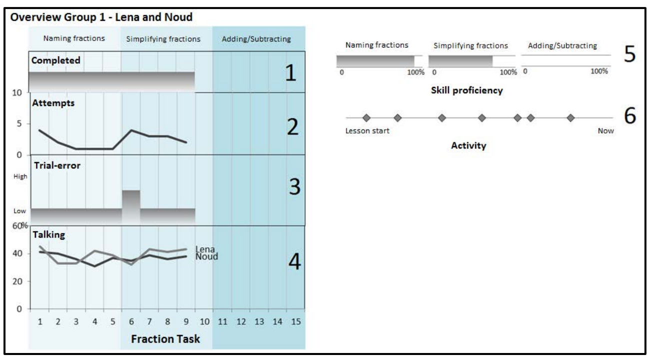

Figure 1: Screenshot of the dashboard with an opened group overview page, with information about six indicators (marked with numbers 1-6).

On the left, a graph with assignment number on the horizontal axis displays indicators number 1-4. The chance of trial-and-error behavior, indicator 3, was based on a dyad's frequency of activity combined with the number of attempts on an assignment (high frequency of activity and high number of attempts indicating a higher chance at trial-and-error behavior). The amount of talk was based on the sound each student's laptop detected, and thus did not offer information about the content of the conversation. On the right page of the group overview, three bars display the dyads' proficiency at the three fraction skills (5), and below that, a timeline shows the dyads' activity from the start of the lesson until now (6). The dyads' proficiency was based on the number of completed assignments in combination with the number of needed attempts for that particular skill. The activity timeline shows a dot whenever one of the group members gives input or clicks a button within MathTutor.

The dashboard situations were designed in such a way that one of the five groups displayed a specific problem. Those problem scenarios were based on literature about the characteristics of successful and less successful collaboration (e.g., [10, 19]). Two situations entailed a cognitive problem (e.g., a dyad being stuck on a problem), two situations included a social problem (e.g., a dyad showing lack of discussion), and two situations showed a combination of a cognitive and social problem. By setting up the values on the six indicators for one of the five groups in a particular way, the problematic situations were created. The remaining four unproblematic groups' values were kept average. Finally, two situations did not include a problematic group, so all five groups showed average values.

The dashboards in the mirroring condition displayed information about the five groups for each of the six indicators (Figure 2, top). The dashboards in the advising condition displayed the same information, but also included a visual cue (an exclamation mark) that denoted the group that was in need of support (Figure 2, bottom). The group overview of the marked group also contained a text box that explained why the dashboard had marked the group as being in need of support.

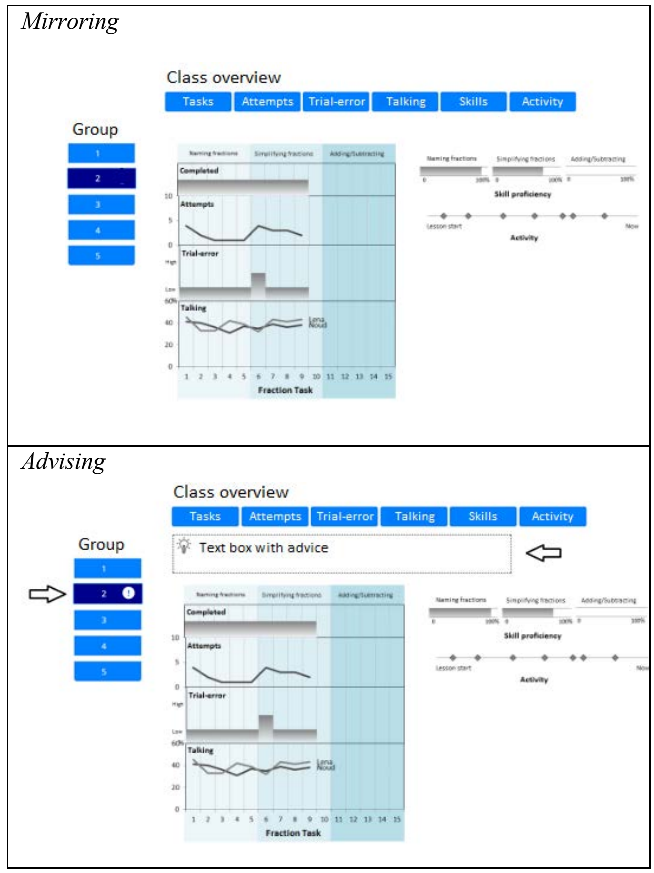

Figure 2: Screenshots of the mirroring and advising dashboards.

The teachers' task was to detect the group that they thought faced a problem (if any). Participants could click all class overview and group overview pages as often as they wanted and could end the situation if they had made a decision by clicking the 'Finish' button. To mimic the classroom situation, in which teachers experience time pressure, the participants had 50 seconds in each situation to decide. If they did not press the 'Finish' button themselves before that time, the situation ended automatically.

After each situation, participants answered a number of questions, namely which group they thought had faced a problem, what type of problem the group faced, whether and how they would intervene in these situations, how much effort it took to answer these questions, and how much confidence participants had in their answer. The amount of effort, which can be regarded as an indicator of experienced cognitive load, was measured with the widely used scale developed by [21], ranging from 1 (very, very little effort) to 9 (very, very much effort). The confidence question was measured on a scale from 1 (very unsure of my answer) to 10 (very sure of my answer).

At the end of the eight situations, participants were asked a number of general questions about the usability of the dashboard and the clarity of the procedure of the experiment. They were also asked to describe their general strategy for interacting with the dashboards.

2.1 Dependent Measures and Analyses

In the present study, two data sources were investigated. First, out of the eight dashboard situations, we extracted the teachers' detection accuracy, meaning the number of situations in which the teachers correctly identified the group that had faced a problem (or correctly identified that there was no problematic group). Furthermore, we calculated the average cognitive load and confidence level associated with selecting a group after each vignette. Because none of the teacher background variables differed significantly between conditions, detection accuracy, cognitive load, and confidence level were compared across the two conditions by means of independent samples t-tests.

Second, we used the automatically generated log-files of the teachers' clickstream while interacting with the dashboard to compare the frequencies of specific actions and to compose teacher-dashboard interaction patterns. The log-files contain chronically ordered information about the specific pages (i.e., class and group overviews) each teacher visited on the dashboard during the 50 seconds that each situation lasted. While interacting with a mirroring dashboard, a teacher had eight options: visiting one of the six indicator overviews, visiting a group overview, and pressing the 'finish' button. The advising dashboard also signaled the problematic group, and therefore had 9 options concerning teachers' actions: visiting one of the six indicator overviews, visiting a "signaled" group overview, visiting a "non-signaled" group overview, and pressing the 'finish' button.

The log-files were reduced in the following two ways. First, because there were two vignettes that did not contain a problematic group, these vignettes were visually different than the others in the advising condition (because there was no alert visible). It was decided to remove the log-files for these two vignettes to obtain a clearer comparison of the strategies teachers used in the two conditions. Second, although participants received an elaborate introduction to the dashboards, it can be assumed that navigating the dashboards was a new experience for most participants. It was therefore decided to remove data from the vignette that was seen first by each participant. As the vignettes were presented in random order, this means for each participant a different vignette was removed from the dataset. Out of the initial 280 vignettes (8 vignettes for all 35 participants), the remaining dataset contained log-files for 91 vignettes for the mirroring condition and 96 for the advising condition.

Frequencies for all possible actions on the dashboard were extracted from the log-files and compared using independent samples t-tests. Furthermore, sequential diagrams were created for the mirroring and advising condition separately that show the transitions between the teachers' actions that were most likely to occur (using the Process Mining Toolkit, see below). As noted, participants were also asked to describe their general strategy for interacting with the dashboards at the end of the experiment. These 35 comments were rather short in nature and thus not systematically analyzed, but we used them to check whether our interpretation of the quantitative results was correct.

As a follow-up step, we zoomed in on each condition and separated the log-files on instances in which teachers accurately selected the problematic group and instances in which they did not. We then created sequential diagrams for these split log-files separately to investigate whether accurate and non-accurate diagnosis were characterized by differing teacher-dashboard interaction patterns.

To perform sequential analysis, the Process Mining Toolkit was used [22], which includes the fuzzy miner algorithm that allows researchers to obtain pattern descriptions in temporally ordered data such as the log-files of teacher-dashboard interaction [1]. The resulting graphs visualize participants' activities (nodes) and the connections between them (edges). We used the settings for the fuzzy minder algorithm following [1]. Only the most important relations between nodes were retained by employing an edge cutoff score, which was set to 0.2. Self-loops were allowed, meaning a relation from an activity to itself was possible. The fuzzy miner algorithm also allows one to set a cutoff score for including nodes. Because we were interested in all eight or nine possible actions on the dashboard, we set the node cutoff score to 0 to retain all actions in the resulting sequential models.

3 Results

3.1 Research question 1: Detection accuracy in mirroring and advising conditions

Out of eight vignettes, participants in the mirroring condition on average identified the problematic group (or lack thereof) correctly 6.65 times (*SD* = 1.06), versus 7.50 in the advising condition (*SD* = 0.62). An independent samples t-test showed this difference was significant, $t(33) = -2.934$, $p = .006$, $d = 0.99$. The average reported cognitive load was significantly higher in the mirroring condition (*M* = 4.52, *SD* =0.97) than in the advising condition (*M* = 2.77, *SD* = 0.89), $t(33) = 5.586$, $p < .001$, $d = 1.89$. The average confidence level associated with selecting a group was significantly higher in the advising condition (*M* = 7.94, *SD* = 1.20) than in the mirroring condition (*M* = 6.74, *SD* = 1.06), $t(33) = -3.118$, $p = .004$, $d = 1.05$).

Thus, the functionality of the dashboard had an effect on teachers' detection of groups that may need attention: in the advising condition, groups were more often detected, it cost teachers less effort, and teachers had more confidence in their decision. Next, we examined how participants interacted with the dashboards.

3.2 Research question 2: Teacher-dashboard interaction patterns

3.2.1 Descriptive values of teachers' actions. Table 1 presents how many actions participants on average performed on each dashboard, how many of those actions concerned group pages and indicator pages, and how many of the five available group pages were visited on average per dashboard situation.

Average per dashboard situation	Mirroring (n = 17)	Advising (n = 18)	p
Total clicks	11.64 (3.02)	9.89 (3.59)	.131
Group page clicks	2.08 (2.48)	4.23 (3.11)	.031
Indicator clicks	7.53 (1.93)	3.65 (3.08)	< .001
Visited groups (1-5)	1.43 (1.68)	3.00 (1.67)	.009

As can be seen, the two conditions did not differ concerning the number of actions, but they did differ concerning the type of pages that were investigated. Participants in the mirroring condition more often looked at the indicator pages (class overviews), whereas participants in the advising condition more often looked at the group overviews and (as a result) on average investigated 3 out of 5 group pages compared to 1.43 out of 5 in the mirroring condition.

3.2.2 Interaction patterns for mirroring and advising condition.. Next, we performed sequential analysis to examine the most often occurring transitions between participants' actions while interacting with the dashboards. We first compared the log-files of all participants in the mirroring condition to those in the advising condition. Figure 3 shows the resulting sequential models. In both models, clicking on one of the six indicators (class overviews) is represented with the six boxes on the right, and clicking on a group overview (alerted or non-alerted) is shown on the left. The two models show considerable similarity. From the start of the dashboard situations, participants either start by visiting the group pages or by visiting the class overviews. In case of the group pages, the self-loops indicate that typically, multiple group overviews are consulted in a row. In case of the class overviews, there are often occurring transitions between the six indicators in the same order in which the indicators are shown on the dashboard (see Figure 2 in the method section).

Some differences between the models can be noticed as well. First, in the advising condition, there is an often occurring transition from the start of the situation to visiting the group overview of the alerted group. This finding means that the participants' attention is probably drawn by the given alert, and participants first look at the alerted page before moving to and between the other group pages. The second difference between the two models are the arrows that go into the 'Stop' state. In the advising condition, there is an arrow from the group page of the alerted group to ending the dashboard situation, and not from the non-alerted pages, meaning that participants were likely to check the alerted group a final time before making a decision about the situation. The other interaction path that led to ending the situation was via the last group overview, the indicator 'Activity'. In the mirroring condition, there are more paths that led to ending the dashboard situation. There is an often occurring transition between visiting a group overview and the stop state, and also between three of the six indicators ('Talking', 'Skills', and 'Activity') and the stop state. This finding means that some participants in the mirroring condition checked these indicators specifically before making their decision about the situation.

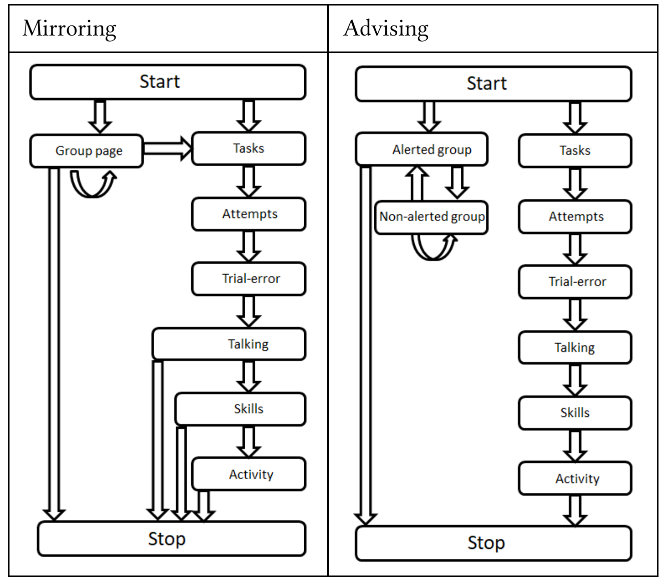

Figure 3: Sequential models for the mirroring (left) and advising (right) condition.

In combination with the frequencies presented in section 3.2.1, it appears that in both conditions the group and class overviews were monitored, but in the advising condition, participants typically focused more on the group overviews and started by reacting to the given alert. The written comments from the participants at the end of the experiment reflect these findings. For example, comments include "Your attention is drawn first by the alarming exclamation mark, but I found it important to look at all groups. I checked all the graphs and then the activity (on the right)." Another example is: "I always looked first whether there was an alert, and I checked it. I also found it important to check how much talking there was, the chance of trial-and-error-behavior, and the progress on the tasks".

In the mirroring condition, the participants more often looked at the class overviews, and they differed in which particular indicator had their most interest. The open comments reflect that participants typically started with the class overviews and then looked at particular groups. There is diversity in which indicators are mentioned. For example: "I first looked whether everyone had progressed equally on the task. If any group was way beyond or behind the others, I found that striking and I paid that group extra attention. I also found it important to check how many attempts groups needed." versus "I mostly looked at the chance at trial-and-error-behavior and the amount of talking."

3.3 Interaction patterns for accurate and non-accurate detection of problematic group

As a follow up step, we split the log-files for both conditions into the instances when the problematic group was identified correctly and instances when it was not. Our aim was to investigate whether accurate and non-accurate diagnosis was characterized by different preceding teacher-dashboard interaction patterns. In the mirroring condition, 13 non-accurate diagnoses were compared to 78 accurate diagnoses. In the advising condition, the dataset only included 3 instances of non-accurate diagnoses, which means there was too few data to split this dataset for follow up investigation. We therefore only report on the mirroring condition.

Figure 4 shows the resulting sequential models for accurate (left) and non-accurate (right) diagnosis in the mirroring condition.

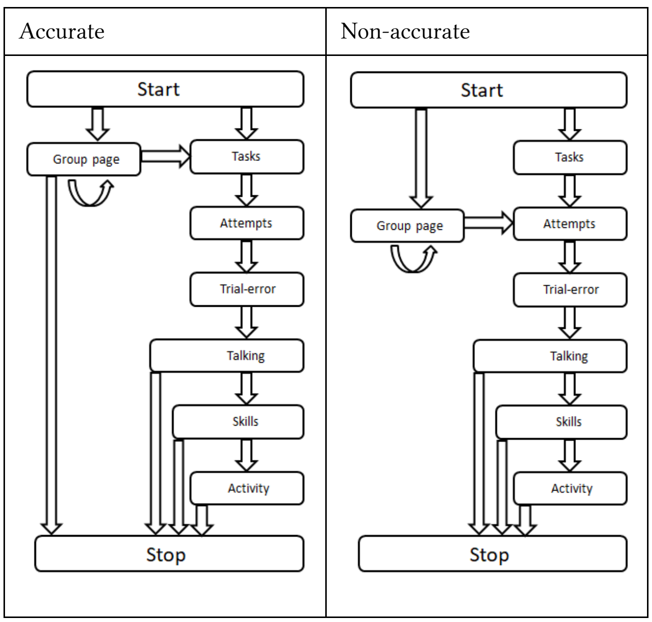

Figure 4: Sequential models for accurate diagnoses (left) and non-accurate diagnosis (right) in the mirroring condition.

The accurate diagnosis model is the same as the model that was generated for the mirroring condition as a whole, see Figure 3 on the left. The model for non-accurate diagnosis differs in two ways. First, there is an often occurring transition between visiting a group overview and looking at the indicator for number of attempts on the task. Second, there is no transitional arrow from visiting a group overview and the Stop state, which is present in the left model. Thus, for the instances with accurate diagnosis, the teacher-dashboard interaction more often ended with teachers looking at one of the group overviews, whereas instances with non-accurate diagnosis were more often preceded by looking at one of the class overviews (indicators). The two models have in common that participants differed in which particular indicator had their most interest, as shown by the multitude of arrows from one of the indicators to the Stop state.

4 Discussion

Teacher dashboards are a specific application of learning analytics: modern technologies that display information about students to inform teachers and to increase teachers' awareness of the situation. Teachers' interaction with a dashboard precedes their interpretation of the situation and thus their subsequent decision making about which group is in need of support. These processes are currently not extensively researched, and in the present study, we combined measures of the accuracy with which teachers detect groups that are in need of support with an investigation into teacher-dashboard interaction patterns. Our aim was to contribute to the field of learning analytics by examining in-depth the interaction between the technology and the user.

Before we discuss the findings of the present study, it must be noted that this study was conducted in a specific context, namely concerning relatively young teachers, small sized groups of collaborating students (i.e., dyads), and the specific domain of mathematics. Each of these characteristics influences the way students interact and how teachers interpret the collaboration. For example, as group size increases, the collaboration between students requires more coordination and teachers might be more likely to focus on this aspect [11]. The study's findings should therefore be considered in light of its specific context, and caution should be exerted when generalizing to other contexts.

The first research question addressed in this study was whether teachers' detection of groups would improve when they interact with an advising dashboard compared to a mirroring dashboard. We indeed found that teachers in the advising condition more often detected the group that was in need of support. Furthermore, it cost teachers less effort, and teachers had more confidence in their decision. These findings confirm our hypothesis and extend earlier research [29] by offering a direct comparison between a mirroring and advising dashboard. Of course, more studies are needed to show whether our findings are robust, but they offer initial indication that the advising dashboard is a form of support that aids the teacher in the crucial task of identifying collaborating groups that may need additional attention. As explained in Section 1, the way a teacher divides his or her attention and moves through the classroom can have a large impact on students' on-task behaviors, and in turn, on their learning outcomes [2, 7].

The second research question concerned what teacher-dashboard interaction patterns emerged when participants interacted with the mirroring and advising dashboard. We found that the interaction patterns showed similarities between conditions, but also some differences. In the mirroring condition, when teachers are given no additional support beyond the provision of information, teachers primarily looked at class overviews and tried to find groups that stood out by means of visual markers. This is in line with earlier studies that show that information about students can be overwhelming, leading teachers to the coping strategy of looking for groups that visible differ from the others, either in positive or negative sense [33]. Another finding for the mirroring condition was that there was more variation in the type of indicator that teachers ascribed most value to and upon which they based their decision about which group needed their support. This finding could point to the importance of teachers' pedagogical beliefs about what aspects of collaborative learning to monitor [9] in relation to how they interact with a dashboard and is an interesting avenue for future research.

In the advising condition, in which teachers were provided with alerts and advice about which group might need support, teachers tended to first look at these alert and subsequently tried to obtain as much information about the other groups as they could to check the dashboard's alert and advice. They thus did not stop monitoring the situation after looking at the alerted group, but continued to process the other available information as well. In the open comments, the teachers indicated they found it important to look at all the groups, and apparently had enough time to do so. The question is what would happen when teachers are under even more time pressure; i.e., whether they would trust the dashboard to point them to the correct groups and would only process information about the group that the dashboard alerted them about.

For the third and last research question, we attempted to determine whether specific teacher-dashboard interaction patterns could be discerned for instances of accurate and non-accurate detection of groups that are in need of support. For the advising condition, there were very few instances of non-accurate detection, and the strategy the participants employed in this condition thus seemed successful. The strategy also proved to be manageable, as teachers reported less mental effort and were able to observe on average three out of five available group overviews. In the mirroring condition, the distinction between the log-files belonging to instances of accurate and non-accurate detection of groups was that for accurate detection, participants more often ended their process of decision making by looking at one of the group overviews instead of one of the class overviews. It could be that because of time pressure, participants did not always have time to look at the group overviews after they were done looking at the class overviews, and thus did not check their initial idea of which group they thought needed support by diagnosing more in-depth. It could also be that they did not even get to the stage of detecting a relevant event in one of the groups, and thus did not go beyond monitoring the information at class level.

In the mirroring condition, there was thus more variation in the extent to which teachers were able to process the information and to detect the groups that were in need of support. In relation to study [9], in which teachers' monitoring competence is measured by having them rate standardized video with collaborative situations, looking at teachers' interaction pattern with the dashboard may be a clue as to whether a teacher needs more advanced support than the mirroring dashboard provides. Thus, a real-time measurement of what support teachers need could be implemented, besides or in conjuncture with an instrument that measures monitoring competency before teachers interact with the dashboard.

Together, the findings of this study show that teachers' detection of groups that may be in need of support is directly influenced by the amount of support a dashboard offers. In the field of teacher dashboards, it is mostly mirroring dashboards that are currently under investigation [24]. The present study shows that even simple visual cues (in the form of alerts) to direct teachers' attention can have considerable effect: not only do they steer teachers' attention towards specific information, but they also lead to less effortful and more confident decision-making by the teachers. Our understanding of these findings is that in the advising condition, because of the given alerts, teachers were supported by steering their attention to a particular group, after which they could spend mental effort on zooming in on information about not only the alerted but also about the other groups. The work by [4] is therefore an interesting link to the present study because it concern the exploration of a dashboard that already narrows down the information shown to the teacher, instead of offering all information available. This is a different strategy of unburdening the teacher in the process of interpretation of classroom information.

In more general terms of teacher-dashboard interaction, the difference between mirroring and advising dashboard seems to be that the way the teacher and the dashboard divide responsibility [25] to support students is balanced differently. In the mirroring condition, the teacher has full responsibility over detecting groups in need and is the one looking for possible outliers. In the advising condition, the system performs the initial assessment of the information about the collaborating students and the teacher checks whether the system's alerts make sense. If the teacher can rely on the system to perform this first step, it relieves the teacher of processing all information and thus saves mental effort. As a result, the effort of deciding of how to move through the classroom and divide attention over the collaborating groups could be smaller for the teacher.

However, it must be noted that the process of detecting a group was simplified in the present study. Because of the chosen methodology of standardized, fictitious situations that were presented to teachers, some elements of the real classroom that could influence detection of groups were left out of this investigation. In particular, in an actual classroom students show behavioral cues to draw the teacher's attention, and the teacher knows the students. Both factors influence teachers' decision making and how the teacher moves through the classroom [6]. On the other hand, by using a structured study as the present one in which all participants are subjected to the same situations, the influence of these complicating factors can be ruled out to perform initial fundamental studies on the effect of different types of dashboards. Several other groups of researchers have employed such a design to investigate teacher-dashboard interaction (e.g., [3, 18, 28]).

Another limitation of the present study is that we did not investigate whether subsequent phases of teachers' decision making are also influenced by the role of the dashboard. After detecting an event, teachers interact with their students and need to decide what support to provide to students, if any [33]. In other words, there are several more steps between a teacher's detection of a group that may be in need of support and the teacher's actual given support that has an effect on student collaboration. The experiment presented here focused on the first step, and at least provides strong indication that the role a dashboard fulfills influences the initial important step of which group teachers approach in the classroom. Whether learning analytics dashboards can also support teachers in the complicated question of which type of intervention is most appropriate for a particular group of students [15] remains an avenue for future research.

ACKNOWLEDGMENTS

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

REFERENCES

[1] M. Bannert, P. Reimann, and C. Sonnenberg (2014). Process mining techniques for analyzing patterns and strategies in students' self-regulated learning. Metacognition and Learning, 9, 161-185. doi:10.1007/s11409-013-9107-6

[2] M. M. Chiu (2004). Adapting teacher interventions to student needs during cooperative learning: how to improve student problem solving and time on-task. American Educational Research Journal, 41(2), 365-399. doi:10.3102/00028312041002365

[3] I.-A. Chounta and N. Avouris (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. doi:10.1007/s10639-014-9355-3

[4] V. Echeverria., R. Martinez-Maldonado, S. Buckingham Shum, K. Chiluiza., R. Granda and C. Conati (2018). Exploratory versus Explanatory Visual Learning Analytics: Driving Teachers' Attention through Educational Data Storytelling. Journal of Learning Analytics, 5(3), 72-97.

[5] R. M. Gillies and M. Boyle (2010). Teachers' reflections on cooperative learning: Issues of implementation. Teaching and Teacher Education, 26(4), 933-940. doi:10.1016/j.tate.2009.10.034

[6] C. Greiffenhagen (2012). Making rounds: The routine work of the teacher during collaborative learning with computers. International Journal of Computer-Supported Collaborative Learning, 7(1), 11-42. doi:10.1007/s11412-011-9134-8

[7] K. Holstein, B. McLaren, and V. Aleven (2017). SPACLE: investigating learning across virtual and physical spaces using spatial replays. Proceedings of the Seventh International Learning Analytics and Knowledge (LAK) Conference, pp. 358-367.

[8] I. Hoogland, K. Schildkamp, F. Van der Kleij, M. Heitink, W. Kippers, B. Veldkamp, B., and A.M. Dijkstra (2016). Prerequisites for data-based decision making in the classroom: Research evidence and practical illustrations. Teaching and Teacher Education, 60, 377- 386. doi:10.1016/j.tate.2016.07.012

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

[10] C. Kaendler, M. Wiedmann, N. Rummel, and H. Spada (2015). Teacher Competencies for the Implementation of Collaborative Learning in the Classroom: a Framework and Research Review. Educational Psychology Review, 27(3), 505-536.

[11] F. Kirschner, P.A. Kirschner, F. Paas, and J. Janssen (2011). Differential effects of problem solving demands on individual and collaborative learning outcomes. Learning and Instruction, 21, 587-599. doi:10.1016/j.learninstruc.2011.01.001

[12] E. Kosba, V. Dimitrova, and R. Boyle (2005). Using student and group models to support teachers in web-based distance education. Proceedings of the International conference on User Modeling 2005, pp. 124-133.

[13] E. Kyndt, E. Raes, B. Lismont, F. Timmers, E. Cascallar, and F. Dochy (2013). A meta-analysis of the effects of face-to-face cooperative learning. Do recent studies falsify or verify earlier findings? Educational Research Review, 10, 133-149. doi:10.1016/j.edurev.2013.02.002

[14] C. Lang, G. Siemens, A.F. Wise, and D. Gasevic. (2017). Handbook of Learning Analytics. Society for Learning Analytics Research.

[15] T.-J. Lin, M. Jadallah, R.C. Anderson, A.R. Baker, K. Nguyen-Jahile, I.-H. Kim et al. (2015). Less is more: teachers' influence during peer collaboration. Journal of Educational Psychology, 107(2), 609-629. doi:10.1037/a0037758

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

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

[18] R. Mazza and V. Dimitrova (2007). CourseVis: A graphical student monitoring tool for supporting instructors in web-based distance courses. International Journal of Human-Computer Studies, 65(2), 125-139. doi:10.1016/j.ijhcs.2006.08.008

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

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

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

[22] Prom (2010), vesion 6.8. Process Mining Group, Eindhoven Technical University, 2010. http://promtools.org

[23] B.B. Schwarz and C.S. Asterhan (2011). E-moderation of synchronous discussions in educational settings: A nascent practice. Journal of the Learning Sciences, 20(3), 395-442. doi:10.1080/10508406.2011.553257

[24] S. Sergis and D.G. Sampson (2017). Teaching and Learning Analytics to Support Teacher Inquiry: A Systematic Literature Review. In A. Peña-Ayala (Ed.), Learning Analytics: Fundaments, Applications, and Trends, pp. 25-63.

[25] A. Soller, A. Martínez, P. Jermann and M. Muehlenbrock (2005). From mirroring to guiding: a review of state of the art technology for supporting collaborative learning. International Journal of Artificial Intelligence in Education, 15(4), 261-290.

[26] A. Van Leeuwen (2015). Learning analytics to support teachers during synchronous CSCL: balancing between overview and overload. Journal of Learning Analytics, 2, 138-162.

[27] A. Van Leeuwen, J. Janssen, G. Erkens, and M. Brekelmans (2013). Teacher interventions in a synchronous, co-located CSCL setting: Analyzing focus, means, and temporality. Computers in Human Behavior, 29(4), 1377-1386.

[28] A. Van Leeuwen, J. Janssen, G. Erkens, and M. Brekelmans (2014). Supporting teachers in guiding collaborating students: Effects of learning analytics in CSCL. Computers & Education, 79, 28-39.

[29] A. Van Leeuwen and N. Rummel (2019). Orchestration tools to support the teacher during student collaboration: a review. Unterrichtswissenschaft, 47(2), 143-158.

[30] A. Van Leeuwen, N. Rummel and T. Van Gog (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. doi:10.1007/s11412-019-09299-x

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

[32] K. Verbert, S. Govaerts, E. Duval, J.L. Santos, F. Van Assche, G. Parra, and J. Klerkx (2014). Learning dashboards: an overview and future research opportunities. Personal and Ubiquitous Computing, 18, 1499-1514. doi:10.1007/s00779-013-0751-2

[33] E. Voyiatzaki and N. Avouris (2014). Support for the teacher in technology-enhanced collaborative classroom. Education and Information Technologies, 19(1), 129-154. doi:10.1007/s10639-012-9203-2

[34] N.M. Webb (2009). The teacher's role in promoting collaborative dialogue in the classroom. The British Journal of Educational Psychology, 79(1), 1-28. doi:10.1348/000709908X380772

[35] A.F. Wise and Y. Yung (2019). Teaching with analytics: Towards a situated model of instructional decision-making. Journal of Learning Analytics, 6(2), 53–69. doi:10.18608/jla.2019.62.4