

Using Innovative Methods to Explore the Potential of an Alerting Dashboard for Science Inquiry

Rachel Dickler¹, Janice Gobert², Michael Sao Pedro³

Abstract

Educational technologies, such as teacher dashboards, are being developed to support teachers' instruction and students' learning. Specifically, dashboards support teachers in providing the just-in-time instruction needed by students in complex contexts such as science inquiry. In this study, we used the Inq-Blotter teacher-alerting dashboard to investigate whether teacher support elicited by the technology influenced students' inquiry performance in a science intelligent tutoring system, Inq-ITS. Results indicated that students' inquiry improved after receiving teachers' help, elicited by the Inq-Blotter alerts. This inquiry improvement was significantly greater than for matched students who did not receive help from the teacher in response to alerts. Epistemic network analyses were then used to investigate the patterns in the discursive supports provided to students by teachers. These analyses revealed significant differences in the types of support that fostered (versus did not foster) student improvement; differences across teachers were also found. Overall, this study used innovative tools and analyses to understand how teachers use this technological genre of alerting dashboards to dynamically support students in science inquiry.

Notes for Practice

- Learning analytics dashboards are a technological genre that provides real-time monitoring to teachers so they can track students' progress and make pedagogical decisions on the fly to best support learning.
- Our dashboard for science guides teachers to support students on inquiry practices by providing alerts with fine-grained data and visualizations in real time.
- This study shows that the dashboard's real-time alerts were effective in helping teachers respond to student difficulties with science inquiry practices and, correspondingly, in promoting student improvement on science practices after receiving teacher help.
- Analyses of the discourse data showed that there were variations in the types of support that resulted in student improvement, with students benefiting most from higher-level support on how to understand and execute science practices, in addition to content support.

Keywords

Teacher dashboard, epistemic network analyses, science inquiry

Submitted: 23/04/20 — **Accepted:** 18/02/21 — **Published:** 12/04/21

Corresponding author ¹Email: rachel.dickler@gse.rutgers.edu Address: Rutgers Graduate School of Education, 10 Seminary Place, New Brunswick, NJ 08901, USA. ORCID ID: <https://orcid.org/0000-0002-9018-4848>

²Email: janice.gobert@gse.rutgers.edu Address: Rutgers Graduate School of Education, 10 Seminary Place, New Brunswick, NJ 08901, USA. And, Email: janice@appendis.com Address: Apprendis, 57 Village Court, Berlin, MA 01503, USA.

³Email: mikesp@appendis.com Address: Apprendis, 57 Village Court, Berlin, MA 01503, USA. ORCID ID: <https://orcid.org/0000-0003-1036-517X>

1. Introduction

Over the past decade, a great deal of work has been done on the development of technologies designed to support teachers' pedagogical practices (Charleer, Klerkx, & Duval, 2014; Roschelle, Martin, Ahn, & Schank, 2017; Sherin, Jacobs, & Philipp, 2011; Verbert et al., 2014). Some researchers are designing innovative tools such as dashboards (Dillenbourg, 2013) — monitoring tools that can provide teachers with real-time data on students' performance within online learning environments — that help teachers identify and support struggling students in classrooms (Holstein, McLaren, & Alevan, 2019; Martinez-Maldonado, Clayphan, Yacef, & Kay, 2015; Molenaar & Knoop-van Campen, 2017; Schwarz et al., 2018; Tissenbaum & Slotta, 2019; van Leeuwen, van Wermeskerken, Erkens, & Rummel, 2017). Without dashboards, teachers would have to

identify the students who are having difficulty in a classroom based solely on behavioural cues such as raised hands (e.g., Sherin et al., 2011), or by manually tracking each student's progress simultaneously. This is especially challenging because some students will not ask for help or do not know that they need help (Aleven, Roll, McLaren, & Koedinger, 2016). Alerting dashboards (Roschelle et al., 2017), in particular, aim to help teachers identify and diagnose who is struggling by providing them with the contextual information they need to support students' learning in real time. These contextual alerts are essential because it can also be challenging to support struggling students, even after they are identified, without an in-depth understanding of their particular difficulties (Shulman, 1987). While some researchers have developed and evaluated dashboards that notify teachers of student completion of activities (e.g., Schwarz et al., 2018; Tissenbaum & Slotta, 2019), dashboards with *contextual* alerts based on automated scoring of student performance are a new technological genre and have only recently been developed for complex domains such as science inquiry. Inquiry is a very important aspect of national and international science frameworks. Thus, it is critical to examine how dashboards that can alert teachers about specific student difficulties with inquiry fit into actual classroom settings (Laurillard, 2008), guiding teacher support of struggling students. Specifically, it is essential that these technologies be implemented and tested in classrooms using methodologies that account for the impact on student learning of teachers' use of the technology (Roschelle et al., 2017). The following sections provide an overview of current teacher dashboards; how science inquiry-alerting dashboards have the potential to meet a critical need in science education; current methods for evaluating dashboard use; and how to build on these methods to better support teachers' pedagogical practices and, in turn, students' inquiry practices.

1.1. Dashboards

Dashboards may be accessed on a variety of platforms (e.g., tablets, computers, interactive tabletops; Holstein et al., 2019; Martinez-Maldonado et al., 2015; Molenaar & Knoop-van Campen, 2017; Schwarz et al., 2018; Tissenbaum & Slotta, 2019; van Leeuwen & Rummel, 2018) and are designed so that teachers can navigate through displays of student performance (e.g., flow charts, histograms, text boxes) at varying levels of specificity (Charleer et al., 2014; West, 2012). The types of data related to student performance differ based on the respective online student environment, but generally dashboards may include information regarding the number of activities students have completed and their overall performance on activities (Verbert et al., 2014).

A number of innovative dashboards have been developed to accompany mathematics environments for students, including Lumilo (Holstein, McLaren, & Aleven, 2018a, 2018b, 2019), Snappet (Molenaar & Knoop-van Campen, 2017, 2018), MTDashboard (Martinez-Maldonado, Kay, Yacef, & Schwendimann, 2012; Martinez-Maldonado, Kay, Yacef, Edbauer, & Dimitriadis, 2013; Martinez-Maldonado et al., 2015), SAGLET (Segal et al., 2017), and TORch (Van Leeuwen & Rummel, 2018). Few dashboards, however, have been developed for science, and these dashboards are not specific to science inquiry. For example, the WISE dashboard (Matuk, Gerard, Lim-Breitbart, & Linn, 2016), the CK Biology curriculum dashboard (Acosta & Slotta, 2018), and SAIL Smart Space's corresponding tablet tool (Tissenbaum & Slotta, 2019) do not provide teachers with any data on students' inquiry practice competencies, such as asking questions/hypothesizing, carrying out investigations/collecting data, or analyzing and interpreting data (as emphasized by policy documents such as the Next Generation Science Standards (NGSS, 2013)). Additionally, these science dashboards either do not provide any alerts (e.g., Matuk et al., 2016; Acosta & Slotta, 2018) or only provide alerts to teachers when students complete an activity (e.g., Tissenbaum & Slotta, 2019), versus real-time alerts to teachers regarding students who are struggling and the specifics of those students' difficulties. Without detailed contextual alerts based on automated assessment of student performance, teachers must conduct their own evaluations of student progress. This process is both challenging and time-consuming, particularly for science inquiry practices (Pruitt, 2014). We argue that alerting dashboards, driven by fine-grained, real-time assessment of students' science inquiry practices, have the potential to provide teachers with the assessment data that is needed to realize the NGSS. In particular, an alerting dashboard can support teachers in changing classroom instruction on the fly and/or providing more targeted feedback to small groups of students or individual students (Tabak & Kyza, 2018). It is important, however, that we understand how this technological innovation is taken up by teachers to examine whether it fits into a "credible educational ecology" (Laurillard, 2008). This research can then also be used to inform future design iterations that better support teachers and, in turn, students.

1.2. Dashboard Evaluation Methods

As mentioned above, a key area of need is understanding and testing whether and how the innovative technological genre of alerting dashboards is useful to teachers and correspondingly to student learning. To date, most dashboards have been pilot tested with and evaluated by teachers through interviews and observations of teachers' interactions with dashboard mock-ups (Matuk et al., 2016), story boards (Molenaar & Knoop-van Campen, 2018), or simulated student data (Holstein et al., 2019; Van Leeuwen & Rummel, 2018). While these techniques are valuable for initially obtaining teacher feedback on the

visualizations and features within the dashboard, they do not allow for understanding of the nuanced details about which features would be used, and how they would be used within an actual classroom setting. Classrooms with 20 or more students are often hectic, time constrained, and unpredictable relative to a laboratory setting. Therefore, it is essential to also conduct implementations of the dashboard within actual classroom settings, as done recently for dashboards in the domains of both mathematics (Holstein et al., 2018a, 2018b, 2019; Molenaar & Knoop-van Campen, 2017) and science (Tissenbaum & Slotta, 2019). In particular, researchers explored factors such as student performance outcomes related to the dashboard implementation (Holstein et al., 2018a, 2019; Tissenbaum & Slotta, 2019), how often teachers accessed the dashboards (Molenaar & Knoop-van Campen, 2017), and how teachers' attention was distributed among students in response to the dashboard (Holstein et al., 2018b; Martinez-Maldonado et al., 2013).

Studies, however, have yet to address the influence of contextual real-time alerts on classroom interactions and subsequent effects on student learning. Additionally, while some studies captured data such as types of teacher actions in response to dashboards (e.g., Molenaar & Knoop-van Campen, 2017) and video of teachers responding to dashboards (Tissenbaum & Slotta, 2019), studies have yet to conduct an in-depth examination of the discursive teacher supports elicited by dashboard use and how the patterns of supports relate to student performance outcomes. The discourse (Gee, 2004), or conversation, that occurs between teachers and students as a result of the technology can reveal a great deal about how teachers support students' understandings in science (Juwon & Creighton, 2005), including in relation to student outcomes (Howe, Hennessy, Mercer, Vrikki, & Wheatley, 2019; Manz & Renga, 2017). This study extends this existing work by studying the discourse elicited by rich alerts in science classrooms. Furthermore, epistemic network analysis (ENA; Shaffer, Collier, & Ruis, 2016), used to examine dynamic interactions in other science education studies (Rupp et al., 2009; Shaffer et al., 2016; Svarovsky, 2011), is used here to examine the differences in patterns of teacher supports in relation to learning outcomes.

In applying these methods to the science inquiry alerting dashboard, Inq-Blotter, this research addresses the gaps in the literature in relation to the lack of contextual alerting dashboards for science inquiry and rigorous testing of alerting dashboards with attention to discourse in science inquiry classrooms. The paper attends to both student learning outcomes (e.g., Do students improve after receiving help from a teacher informed by dashboard analytics? Do students maintain that improvement over time? How does this improvement compare to students who were not helped by the teacher?) and patterns in teachers' discursive supports informed by the dashboard (e.g., How do teachers respond to dashboard alerts? What types of discursive supports are provided by teachers based on dashboard alerts?). Specifically, ENA is applied to understand how student improvement in the intelligent tutoring system, Inq-ITS, relates to the patterns of supports provided by teachers in response to alerts from the corresponding science inquiry dashboard, Inq-Blotter.

2. Inq-ITS and Inq-Blotter

The Inq-ITS environment was developed using evidence-centred design (ECD; Mislevy, Behrens, Dicerbo, & Levy, 2012) to support middle school students in carrying out inquiry investigations aligned to the NGSS (2013) in the domains of life, physical, and earth sciences. As students progress through interactive inquiry stages (i.e., asking questions/forming hypotheses, carrying out investigations using realistic simulations/collecting data, and analyzing and interpreting data collected using the simulations), all of their actions are logged and automatically scored using knowledge-engineered and educational data-mined algorithms (Gobert, Sao Pedro, Raziuddin, & Baker, 2013; Gobert, Baker, & Sao Pedro, 2014; Gobert & Sao Pedro, 2017; Gobert, Moussavi, Li, Sao Pedro, & Dickler, 2018). This previously validated automated (Gobert et al., 2013) scoring is used to trigger alerts to teachers on students' difficulties in the recently developed alerting dashboard Inq-Blotter (Gobert et al., 2018). As a result, Inq-Blotter is the only dashboard, to our knowledge, that provides alerts based on rich, authentic assessment of students' inquiry competencies. Alerts provide teachers with verbal descriptions and visualizations regarding student performance, including explanations of student difficulties, progress bars with scores on each of the inquiry practices, and information on the activities students completed. The inquiry alerts occur at both the individual level (e.g., a student is struggling with analyzing data) and the class level (e.g., 60% of students are struggling with analyzing data; see Materials section for further details). The initial Inq-Blotter alerts were developed based on both classroom observations and teacher interview data. The alerts were iteratively reviewed and tested by teachers on paper before being implemented in the online dashboard system (Gobert et al., 2018). A randomized controlled study comparing the effects of teacher support elicited by the Inq-Blotter dashboard with alerts versus without alerts (i.e., no performance data) found that students supported by a teacher using the former improved more than students helped by a teacher using the latter (Sao Pedro, Gobert, & Dickler, 2019). It is important, however, to more closely examine *how* teachers are using the alerting dashboard in their classrooms and how this facilitates student learning. Specifically, teacher-student discourse is essential for understanding how the dashboard impacts interactions in the classroom and correspondingly fosters student learning gains.

In this study, we examined teachers' use of the Inq-Blotter alerting dashboard and the corresponding association with student performance. Specifically, we examined whether this intervention improved students' performance on the inquiry practice on which they were helped in their next activities and whether this improvement was robust. We then used ENA to identify differences in the teachers' discursive support patterns in relation to student improvement as well as to identify differences in the support patterns between teachers.

3. Methods

3.1. Participants

Participants included two grade 8 teachers, Ms. A and Mr. B, and their students. Data collection took place over a single class session/period for each class section. Ms. A taught at a southern US middle school (16% of students at the school received free or reduced lunch) and had five class sections with 22 to 30 students in each section (a total of 147 students). Mr. B taught at a northwestern US middle school (57% of students at the school received free or reduced lunch) and had six class sections with 23 to 31 students in each section (a total of 166 students). The participating teachers used Inq-Blotter while their students completed virtual labs in Inq-ITS during their regular science class periods. Teachers and students continued to regularly use the student and teacher platforms in their science classes following the completion of the study. A total of 35 students from Ms. A's class and 56 students from Mr. B's class were helped in response to Inq-Blotter during the data collection sessions. Only 20 students from each teacher (40 students in total) went on to complete a second activity in Inq-ITS after being helped by the teacher. Only 35 of those 40 students went on to complete a third activity. We used the data from students who completed two ($N = 40$ students) or three ($N = 35$ students) activities in this study to examine whether students improved on an inquiry practice in future activities after being helped by the teacher on the practice and whether this improvement was maintained across multiple activities. These students did not receive any help from the teacher during their second or third activity. Additionally, a comparison group of 35 students who completed at least three activities but were *not helped by the teacher at any point* during the study was randomly selected based on matched criteria to the 35 students who were helped in response to an Inq-Blotter alert (see Analyses section for further details).

3.2. Materials

Inq-ITS (Gobert et al., 2013; Gobert & Sao Pedro, 2017) consists of virtual lab activities for middle school in the domains of life, earth, and physical science that are aligned to the NGSS (2013). Each virtual lab activity contains four stages: students ask questions/form hypotheses about a scientific phenomenon; then carry out investigations/collect data using a simulation; then analyze and interpret their data; and finally write an explanation based on their findings in the claim, evidence, and reasoning format (see Figure 1). Students' actions in the first three stages are logged and used for automated assessment of inquiry practice competencies (i.e., question forming/hypothesizing, collecting data, and analyzing and interpreting data; see Measures section for further details). Students from Ms. A's and Mr. B's classes completed virtual lab activities in a variety of topics (flower growth, collisions, sound waves, lunar phases, or phase change). Each topic has three or four lab activities where students investigate different goals. For example, in the phase change lab, students complete three activities to investigate the following three goals: (1) determine how the amount of heat affects the boiling point of water, (2) determine how the amount of ice affects the melting point of ice, and (3) determine how the size of a container affects the boiling point of water.

Inq-Blotter (Gobert, Sao Pedro, Betts, & Baker, 2016; Gobert et al., 2018; Gobert & Sao Pedro, 2017) provides real-time alerts to teachers on students' inquiry competencies for NGSS practices based on the automated assessment in Inq-ITS. Teachers receive alerts on the centre panel of the dashboard (see Figure 2) any time students are automatically assessed as performing below a proficiency threshold for a practice in Inq-ITS (Gobert et al., 2018). Teachers can organize the alerts by the type of inquiry practice students are having difficulty with (asking questions/hypothesizing, collecting data, etc.) or by recency (based on how long the students have been struggling). Teachers can also view a student list with all active students displayed alphabetically. Several settings allow teachers to control the number of alerts displayed and the thresholds for when alerts are displayed. The alerts contain diagnostic information on the specific difficulty a student is having with a practice (e.g., the student is struggling with hypothesizing in terms of understanding the concept of an independent variable), the amount of an activity that the student completed (e.g., the student is still in the first stage of an activity), and other prior difficulties the student had while completing the activity (e.g., the student previously had difficulties with forming a testable hypothesis; see the right panel of Figure 2).

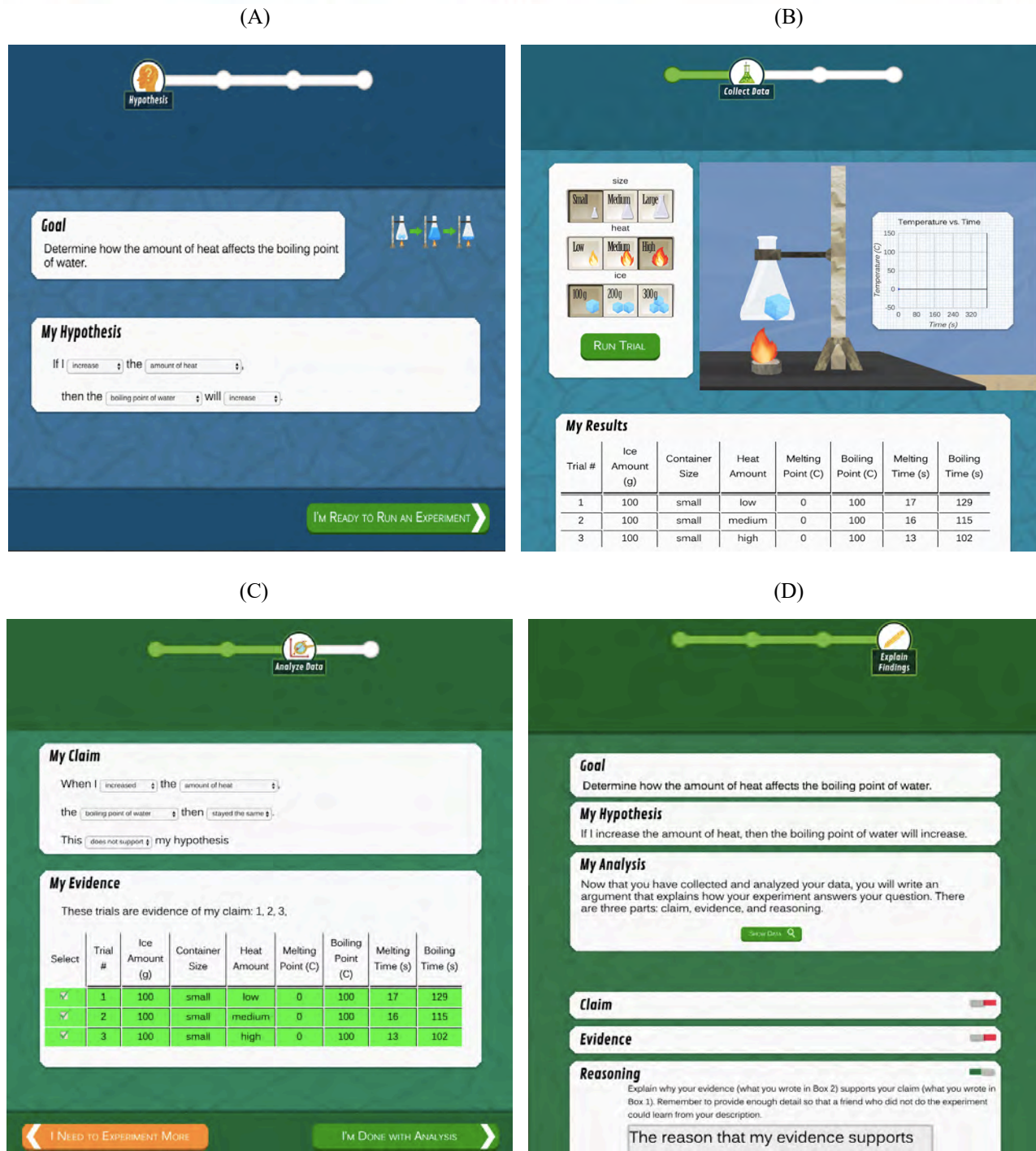


Figure 1. Screenshots of the Inq-ITS forming questions/hypothesizing (A), collecting data (B), analyzing and interpreting data (C), and explaining findings (D) stages from the Phase Change virtual lab.

For our research purposes, the dashboard also contained a microphone button (see the right panel of Figure 2) that the teacher was instructed to press to start a recording when helping a student in response to an alert. Voice recordings were captured using the internal microphones on the teachers' devices (tablets) and were automatically triangulated with the log file data from both Inq-ITS and Inq-Blotter through our systems' infrastructure based on time stamps.

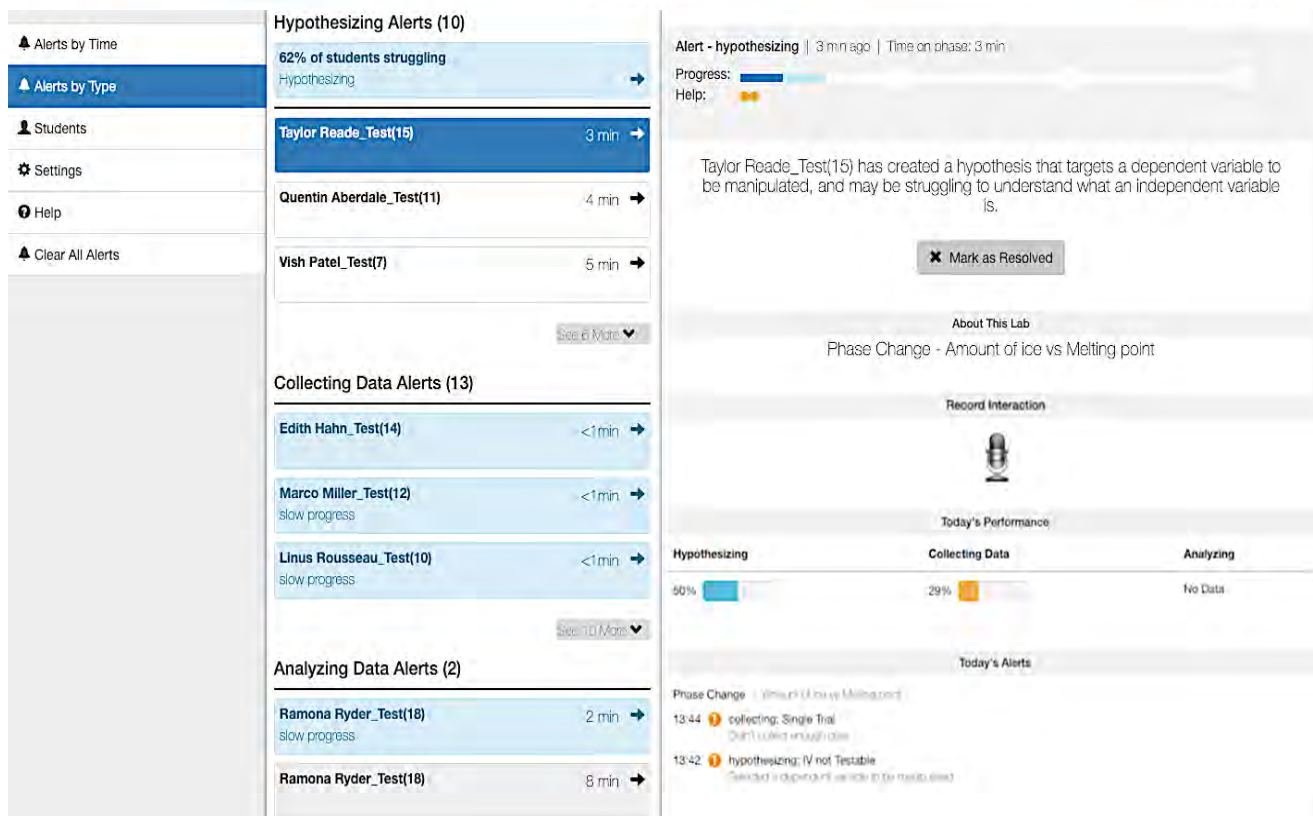


Figure 2. Screenshot of an Inq-Blotter alert with the microphone voice recording feature; clicking on a student’s name presents the teacher with diagnostic information about the student’s difficulty.

3.3. Procedure

Data were collected during the teachers’ and students’ regular science class periods. The researcher explained the capabilities and functionality of Inq-Blotter to the participating teachers before the start of the study. The researcher also explained that the microphone button was to be used any time the teacher helped a student based on information in the dashboard. Inq-Blotter was accessed by teachers through a browser on a tablet device. Teachers had an opportunity to ask questions about any of the information or visualizations within the dashboard. Students were introduced to Inq-ITS by their teacher and then began completing a virtual lab activity on their computers. The specific lab activity topics were determined by the teachers. While students completed the virtual lab activities in Inq-ITS, teachers received real-time alerts in Inq-Blotter on their tablet devices. Any time the teacher responded to an alert, they would press the microphone button to record the discourse and press the button again to end the recording. Data collection lasted a single class period for each class section. Not all students were helped over the course of the class period, and the students completed between one and four virtual lab activities. Ms. A helped 35 students and Mr. B helped 56 students based on the alerts in the dashboard (91 students in total). Of this total number, only 20 students from each teacher went on to complete at least one more activity in Inq-ITS after being helped (40 students in total). A total of 35 of those 40 students went on to complete at least two activities after being helped as part of this study. These students did not receive help from the teacher during either their second or their third activity. Teachers and students had continued access to the Inq-Blotter and Inq-ITS technologies following the completion of the study for use during their regular science class periods.

3.4. Measures

In this study, we tested whether students improved on inquiry practices (on which they were helped) after being helped by a teacher using Inq-Blotter. Performance on inquiry practices was determined using the automated scoring algorithms in Inq-ITS (Gobert et al., 2013, 2014). The automated scoring is based on knowledge-engineered and educational data-mined algorithms validated in prior studies (Gobert et al., 2013, 2018; Gobert & Sao Pedro, 2017). These algorithms are applied to students’ interactions to capture fine-grained student performance in real time on each of the practices in the first three stages of the virtual lab activities: asking questions/hypothesizing, carrying out investigations/collecting data, and analyzing and

interpreting data. Hypothesizing is scored based on whether students identify an appropriate independent variable and dependent variable, and on whether the variables selected are aligned with the goal of the activity. Collecting data is scored based on whether students run controlled trials and whether those trials target the appropriate independent variable. Analyzing and interpreting data is scored based on whether students make a claim that reflects the results of their investigation, identify whether their findings support their initial hypothesis, and back up their claim with sufficient and appropriate evidence. The automated scoring algorithms assign students a score between zero and one points (i.e., the average of the components listed above for each practice that are scored as binary) for each practice within each activity. These scores are passed to Inq-Blotter in real time to trigger alerts. These scores were used in this study along with additional log data from Inq-ITS to identify how many activities students completed and the topic of the virtual lab activities.

All teacher actions in Inq-Blotter (opening an alert, resolving an alert, voice recordings) were automatically logged and time-stamped. These data points were extracted from the database to determine when and how the alerts were used by the teachers. In particular, data on which alerts were accessed, the time at which the alerts were accessed, and the content of alerts were pulled from Inq-Blotter. These data were triangulated with the students’ logs to synchronize their performance before and after they were helped by a teacher.

Table 1. Types of discursive support given by teachers with examples (Dickler et al., 2019b)

Support type	Definition	Example
Orienting scaffold	Directing attention to a particular practice	“Let’s look at analyzing your data... Let’s look at your data.”
Conceptual scaffold	Definition/explanation of an inquiry practice	“The independent variable is what’s manipulated. So what are you changing here?”
Procedural scaffold	Information on the steps involved in an inquiry practice	“You always run a control[led trial] and you always run it more than one time. Why do you think you need to run it more than one time?”
Instrumental scaffold	The exact actions to take to complete the inquiry practice	“It’s saying to you that it didn’t change... So you need to click ‘no change.’”
Content comment	A statement regarding scientific domain-related content	“So that’s saying, ya know, if you have a shorter string the wave speed will be faster; if you have a longer string it’ll be faster.”
Content question	Asking the student about scientific domain-related content	“So you’re gonna change the speed of the wave and that’s gonna affect the loudness?”
Evaluative comment	Statements regarding whether student work is correct or incorrect	“According to this you’re struggling with your hypothesis.”

A total of 91 voice recordings of teacher support in response to Inq-Blotter alerts were also extracted. Only 40 of the recordings (with the 40 students who went on to complete at least one additional activity after being helped by the teacher) were used in this study. These recordings were identified using the triangulated Inq-ITS and Inq-Blotter log data. The voice recordings were transcribed by two researchers and split into clauses according to the speaker (i.e., teacher or student). The teacher clauses ($N = 190$) were then coded for the types of teacher supports provided to the students. The categories for types of teacher supports were determined based on a previously developed taxonomy for teacher inquiry scaffolds (Dickler, Gobert, & Sao Pedro, 2019a; Dickler, Li, & Gobert, 2019b; Gobert et al., 2013), as well as categories of codes used by researchers to identify the general types of support provided by teachers (e.g., Manz & Renga, 2017). Specifically, the types of supports coded in this study included orienting (the teacher directs the student’s attention to a particular component of a step/task), conceptual (the teacher explains an inquiry practice and the purpose of the practice), procedural (the teacher provides the student with general information about how to carry out the inquiry practice), instrumental (the teacher tells the student exactly what to do to complete the inquiry practice), content comment (the teacher tells the student about scientific facts or principles related to the topic of the investigation), content question (the teacher asks the student about scientific facts or principles related to the topic of the investigation), and evaluative comment (the teacher tells the student whether their work is correct or not); see Table 1 for examples. In terms of the taxonomy, the first four types of supports are scaffolds specifically for inquiry practices. The lower-level orienting and instrumental inquiry supports are intended to help the student focus and move through

the lab, whereas higher-level inquiry supports such as conceptual and procedural supports involve helping students to more deeply understand the inquiry practices. Content comments, content questions, and evaluative comments do not directly address inquiry practices.

Individual clauses of the voice recordings were coded by two researchers for whether they fit any of the seven codes. The researchers were first trained on the coding scheme for types of teacher supports using recordings not included in this study and then scored 100 teacher clauses from this study independently. Inter-rater agreement was calculated and the raters were found to agree on 97% of assigned codes. The disagreements were discussed and the final agreed-upon codes were used for analyses. One rater coded the remaining clauses ($N = 90$) independently. These coded clauses were necessary to build the epistemic networks, as explained in the Analyses section.

3.5. Analyses

Using the triangulated data from Inq-ITS and Inq-Blotter, we first identified the practices on which students were helped by the teacher, student performance before being helped (i.e., their first opportunity in their first virtual lab activity), and student performance on the practice on which they were helped in the next Inq-ITS virtual lab activity they completed (i.e., their second opportunity in their second virtual lab activity). We also identified a subset of students who went on to complete a third virtual lab activity ($N = 35$ of the 40 students) and their performance on their third activity (i.e., their third opportunity). Data from the Inq-Blotter log files were used to confirm that teachers did *not* access or provide support to these students based on alerts on either their second or their third opportunity. Student performance data was examined for *only* the practice of interest (i.e., the practice on which the student had difficulty and was correspondingly helped with based on the dashboard alert). Analyses were not separated by the practice of difficulty (e.g., only students helped on hypothesizing/question formation, only students helped on collecting data, etc.) due to the small number of data when broken down by individual practice. Therefore, the dependent variable was student improvement on the practice of difficulty (though the specific inquiry practice differed between students).

Additionally, for each of the 35 students helped by a teacher in response to an Inq-Blotter alert, we randomly identified a matched student based on the following criteria: the student was never helped by the teacher based on an Inq-Blotter alert, the student had the same teacher (either Ms. A or Mr. B), the student received the same score on the inquiry practice of difficulty in the same activity (as the helped student before being supported by the teacher), and the student went on to complete at least two additional activities (three activities in total). It was important to match students based on these criteria to eliminate other possible explanatory factors for differences in performance. We could then track how students who were never helped on the practice (with the same initial performance on the practice of difficulty as students who were helped) progressed over time. Alerts did appear for these students in the dashboard but were not responded to by the teacher for multiple possible reasons, including the teacher not having sufficient time during the class period to respond to all alerts or the student moving on to a new activity (and the alert disappearing from the dashboard) before the teacher was able to respond to the alert.

In terms of the statistical analyses for the data from the 40 students who were helped by a teacher in response to an Inq-Blotter alert and went on to complete a second activity, descriptive statistics were used to identify the proportion of students who improved on an activity (i.e., their score on the inquiry practice increased from their first opportunity before being helped to their next opportunity) or did not improve (i.e., their score on the inquiry practice stayed the same or decreased from their first opportunity before being helped to their next opportunity). In order to investigate whether students who were helped by their teacher significantly improved from their first to their second opportunity on a practice after being helped by a teacher using Inq-Blotter, a paired-samples t-test was used with an alpha of 0.05.

Additionally, for the subset of 35 of the 40 students helped by a teacher in response to the dashboard (who completed three activities), we used a repeated measures analysis of variance (ANOVA) with an alpha of 0.05 to identify whether student performance on the practice on which they were helped significantly improved across the second and third activities. Paired-samples t-tests were used for follow-up analyses in order to determine whether students significantly improved between the first and second, second and third, and/or first and third activities with a corrected alpha of 0.025.

It was then important to understand how the performance of students who were helped by the teacher and completed three activities compared to that of the struggling students who did not receive any help from the teacher in response to dashboard alerts (i.e., for the 35 students in the matched comparison group). To compare the performance of these groups of students, a repeated measures ANOVA was used with the type of help (i.e., teacher dashboard help versus no help) as a between-subjects factor with an alpha of 0.05. The interaction effect and main effects were explored. The corresponding post hoc comparisons were examined using a corrected alpha of 0.025.

We then explored patterns in teacher support in relation to student improvement using ENA (Shaffer et al., 2016) for the 35 students who completed three activities. The epistemic networks were created using the online graphical interface (Marquart, Hinojosa, Swiecki, Eagan, & Shaffer, 2018) such that the nodes represented the teacher support types, and the

strength of connections between the nodes represented the frequency with which two types of supports co-occurred (see Figure 4, where the intersections between lines in the network are the nodes that represent the type of teacher supports and the lines between nodes are the connections representing frequency of co-occurrence). The stanza size, or number of teacher turns selected, for creating the networks was each instance that a teacher helped a student (i.e., each individual recording), and the individual teacher turns were used as the segments in the networks. Networks were then quantitatively compared based on the mean centroid values using t-tests. The mean centroid values of the networks represented how each network was weighted differently depending on the strengths of relationships between teacher support types. Additionally, networks were qualitatively compared using the visual strengths of connections in each network. Comparisons were also made using the subtracted network (where one network was subtracted from the other to identify the greatest differences in strengths of connections between nodes). In terms of the comparisons, we first used ENA to compare patterns of support based on whether the $N = 35$ students who were helped by teachers using Inq-Blotter correspondingly improved (their score on the inquiry practice increased) or not (their score on the inquiry practice stayed the same or decreased) from their first to their second opportunity to complete the inquiry practice on which they were helped. We then used ENA to examine differences in teacher support patterns that resulted in student improvement from their first to their third opportunity or not. Lastly, we applied ENA to explore whether differences also emerged between the two teachers in terms of how they supported their students (20 students for each teacher; $N = 40$ students), regardless of student improvement.

4. Results

We used descriptive statistics to examine the proportion of $N = 40$ students who were helped by a teacher using Inq-Blotter that improved on the practice on which they were helped from their first to their second activity. We then used further analyses (t-tests) to explore whether the improvement was significant. Descriptive analyses indicated that a majority of students (80%; 32 out of 40 students) improved from their first opportunity (before being helped by the teacher in response to Inq-Blotter) to their second opportunity (after being helped) on the practice. Specifically, 16 out of 20 students from each teacher (80%) improved from their first to their second opportunity. A paired-samples t-test was used to see whether student improvement was significant. Results of the paired-samples t-test revealed that students' improvement was significant from their first opportunity ($M = 0.49$, $SD = 0.21$) to their second opportunity ($M = 0.82$, $SD = 0.27$), $t(39) = -5.119$, $p < 0.001$, $d = 1.32$. These results therefore indicate that not only did a majority of students improve, but also this improvement was substantial (i.e., students' performance increased from an average of 50% on a practice to an average of 83% on the practice out of a possible 100%).

Next we explored whether student improvement was robust across multiple activities for the 35 of 40 total students who completed three activities. Descriptive analyses revealed that a majority of students improved from their first opportunity to their second (28 out of 35 students; 80% of students) and third (27 out of 35 students; 77% of students) opportunity after being helped. The significance of this improvement was examined using a repeated measures ANOVA. Results revealed a significant main effect of the number of activities/opportunities, $F(2, 68) = 22.91$, $p < 0.001$, $\eta^2 = 0.40$. This finding indicated that there were significant differences in student performance on inquiry practices between the first, second, and third activities. Follow-up comparisons using paired-samples t-tests revealed that students significantly improved from their first opportunity ($M = 0.49$, $SD = 0.20$) to their second opportunity ($M = 0.83$, $SD = 0.24$), $t(34) = -5.56$, $p < 0.001$, $d = -0.94$, as was expected based on the initial analyses using paired-samples t-tests. Importantly, students also significantly improved from their first opportunity ($M = 0.49$, $SD = 0.20$) to their third opportunity ($M = 0.85$, $SD = 0.26$; $t(34) = -5.84$, $p < 0.001$, $d = -0.99$). There were no significant differences between students' performance on their second ($M = 0.83$, $SD = 0.24$) and their third opportunity ($M = 0.85$, $SD = 0.26$; $t(34) = -0.49$, $p = 0.63$, $d = -0.08$), indicating that students primarily maintained their improved performance after being helped (with slightly higher performance on their third opportunity; see Figure 3). The fact that this score was maintained (and even slightly increased) from the second to the third opportunity is evidence that the teacher support was robust at supporting students' inquiry learning (i.e., for the majority of students, performance did not dip down on the third opportunity; see Figure 3). It is important to note, however, that some students (who improved from the first to the second activity) did experience a drop in performance from the second to the third activity ($N = 4$), while other students who had not improved from the first to the second activity did experience improvement from the first to the third activity ($N = 3$).

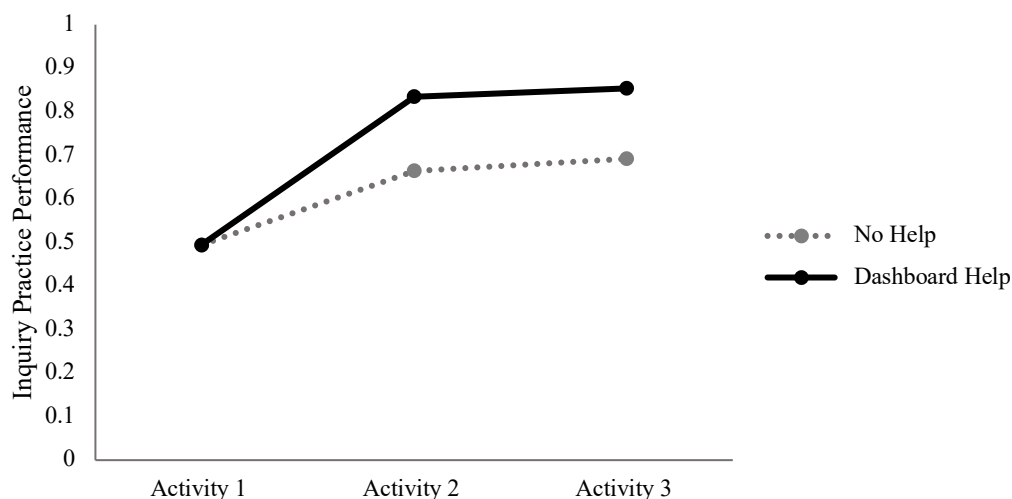


Figure 3. Inquiry practice performance for students who completed three activities after receiving teacher help in response to Inq-Blotter alerts or no help.

We then examined whether this improvement across activities was significant in comparison to students who were not helped by the teacher. Specifically, we compared the performance of the 35 students who completed three activities and were helped by the teacher *relative to* the performance of an additional 35 students who completed three activities but were not helped by the teacher based on an alert in Inq-Blotter. A repeated measures ANOVA was used with type of support as the between-subjects factor (i.e., dashboard help or no help) and performance on the inquiry practice of difficulty on each of the three activities as the within-subjects factor. Results indicated that there was a marginally significant interaction between the type of support students received and the number of activities completed, $F(2, 136) = 2.60, p = 0.078, \eta^2 = 0.04$ (with an alpha of 0.05). Post hoc comparisons revealed that while student performance on activity 1 was the same across groups (because students were matched based on their performance on activity 1; $M = 0.49, SD = 0.20$), students who were helped by a teacher in response to Inq-Blotter performed significantly better on their second activity ($M = 0.83, SD = 0.24$) than students who were not helped ($M = 0.66, SD = 0.35; F(1, 68) = 5.60, p = 0.021$) as well as marginally significantly better on their third activity ($M = 0.85, SD = 0.26$) than students who were not helped ($M = 0.69, SD = 0.36; F(1, 68) = 4.70, p = 0.034$; with a corrected alpha of 0.025). These results are exciting because they indicate that while both groups of students improved on difficult inquiry practices over time, students who received teacher help elicited by an alert improved at a greater rate on their second and third activities (see Figure 3). There was also a significant between-subjects main effect of condition, where students who received help ($M = 0.72, SD = 0.23$) performed significantly better overall than students who did not receive help ($M = 0.62, SD = 0.29; F(1, 68) = 6.28, p = 0.015, \eta^2 = 0.09$). A significant main effect was also found for task ($F(2, 136) = 26.89, p < 0.001, \eta^2 = 0.28$), though this was not explored further due to space limitations.

These findings illustrate that students who received teacher support elicited by the alerts from Inq-Blotter applied their practice competencies in their next activity and maintained that performance better than students who did not receive any support from teachers based on Inq-Blotter alerts. These findings are supported by results in prior studies (Sao Pedro et al., 2019) that also found teacher support elicited by Inq-Blotter to be more effective than supports from teachers without access to a dashboard. While it is exciting to see this robust improvement maintained across activities, it is essential to consider what teachers are saying in response to Inq-Blotter alerts that may foster this improvement. The following analyses used ENA to explore the differences in patterns in teacher support elicited by Inq-Blotter.

4.1. Epistemic Networks by Improvement

First, ENA (Shaffer et al., 2016) was conducted to compare teacher support patterns related to whether students improved or not from their first to their second opportunity after being helped. The voice transcriptions (coded for the types of teacher supports: orienting scaffolds, conceptual scaffolds, procedural scaffolds, instrumental scaffolds, content comments, content questions, and evaluative comments; see Table 1) were grouped according to whether the 35 students (who completed at least three activities) improved from the first activity to the second activity or not. The ENA revealed a significant difference in the patterns of teacher support when students improved ($M = 0.11, SD = 0.49$; left of Figure 4) versus did not improve ($M = -0.46, SD = 0.56; t(34) = 2.45, p = 0.04, d = 1.12$; right of Figure 4) when comparing the mean centroid values (i.e., the values computed based on the weighted network connections in space; Shaffer et al., 2016).

Furthermore, the subtracted network revealed that students who improved on their second opportunity were those who received mostly inquiry scaffolds (including lower-level inquiry support such as orienting and procedural scaffolds) in combination with evaluative comments (i.e., 61% of support associated with improvement; see Figure 5). For instance, Mr. B received an alert in Inq-Blotter that a student was running trials that did not target testing the variables specified in the student’s hypothesis. The supports Mr. B gave to the student based on the alert were the following:

Mr. B: You’re doing an experiment but you’re not targeting your hypothesis [Evaluative Comment], so what’s your hypothesis? [Orienting Scaffold]

Student: [No Response]

Mr. B: ...So what are you changing? [Orienting Scaffold] What is the only thing on here that you’re going to change? [Procedural Scaffold]

This student continued on in the next activities to run investigations that targeted the variables in the hypothesis. However, students who did not improve on their next opportunity received primarily content-related support (both comments and questions) as well as low-level orienting hints (i.e., 68% of support associated with no improvement; see Figure 5). For example, Mr. B received an alert in Inq-Blotter for a student who was struggling to interpret their results about whether the length of a guitar string increases or decreases the loudness of the musical sound. Mr. B provided the following supports:

Mr. B: Are you going to choose increase or decrease? [Orienting Scaffold]

Student: Decrease?

Mr. B: Decrease means less. [Content comment] Okay, then what’s going to happen? [Orienting Scaffold]

Student: Increase?

Mr. B: Increase the loudness [Content comment], right! [Evaluative comment]

This student went on to struggle with interpreting data in their next activities even after receiving this support from Mr. B. It is likely that this support was insufficient to help the student improve in future activities because it was specific to the context of the activity (i.e., whether loudness increased or decreased with the length of a guitar string).

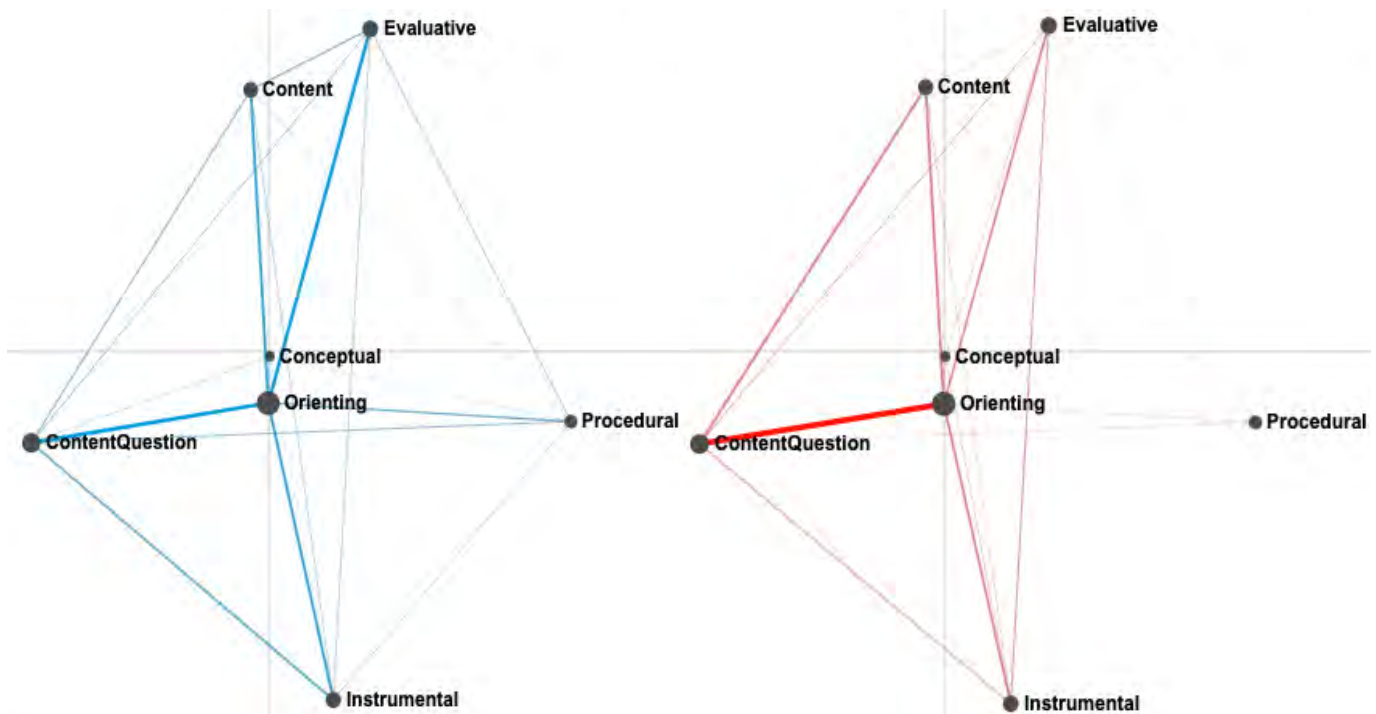


Figure 4. Networks for when students improved (blue; left) or did not improve (red; right) from the first to the second opportunity.

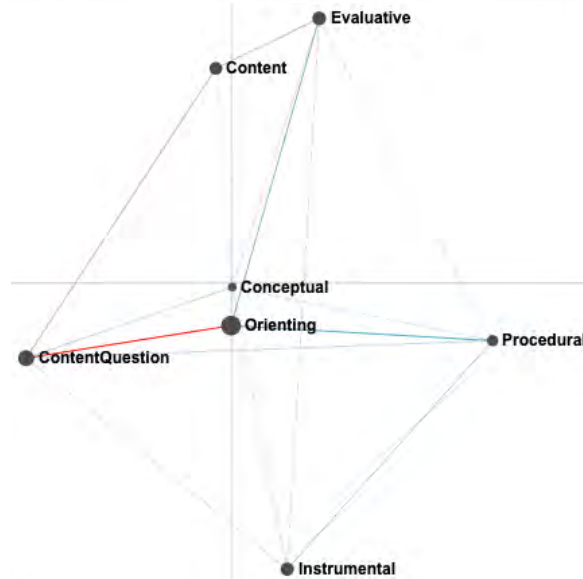


Figure 5. The subtracted network comparing when students improved (blue) or not (red) from the first to the second opportunity.

We also conducted an ENA to compare the types of support related to whether students improved (see left of Figure 6) or not (see right of Figure 6) from their first to their third opportunity. It is important to note that the students who did not improve from the first to the third activity differed from students who did not improve from the first to the second activity (i.e., some of these students improved their performance on their second opportunity but did not maintain this improvement on their third opportunity; $N = 4$), so there is a different subset of students in each condition. Again, we found that there was a significant difference in the patterns of support when students improved ($M = -0.20, SD = 0.54$) versus did not improve ($M = 0.66, SD = 0.84; t(34) = 2.73, p = 0.02, d = 1.40$). Specifically, the strength of connections in the subtracted network revealed that teacher support involving a combination of inquiry (both high-level conceptual and low-level orienting scaffolds) and content-related support was more strongly associated with student improvement (i.e., 88% of support associated with improvement), whereas evaluative comments in combination with low-level orienting and instrumental support did not appear to relate to student improvement (i.e., 68% of support associated with no improvement; see Figure 7).

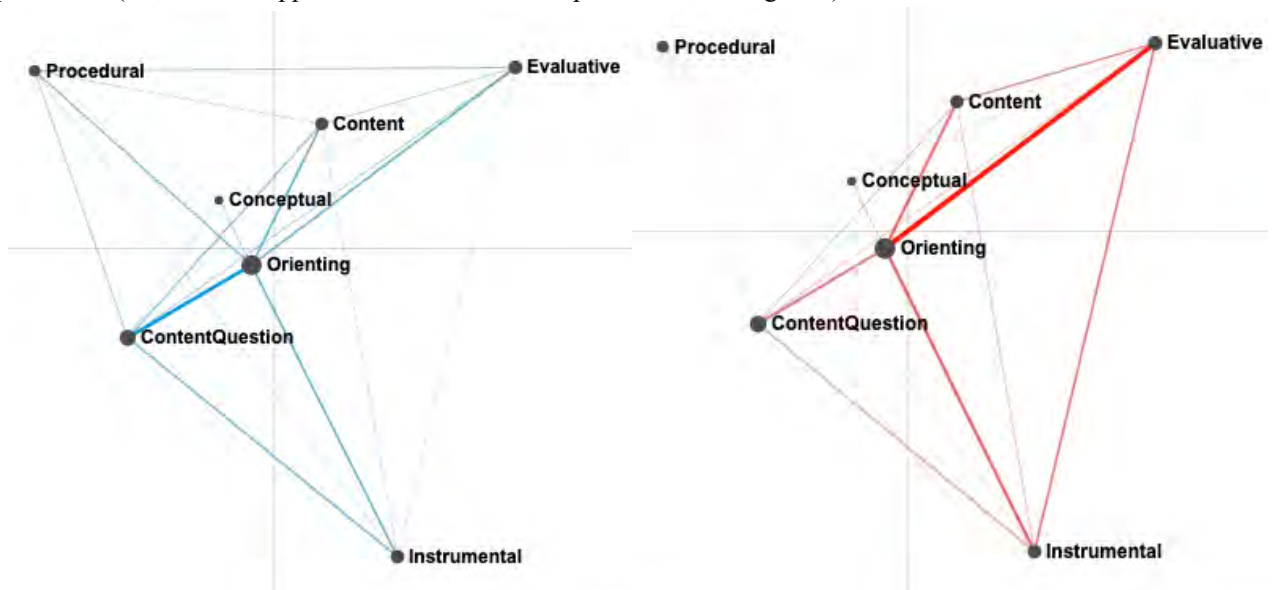


Figure 6. Networks for when students improved (blue; left) or did not improve (red; right) from the first to the third opportunity.

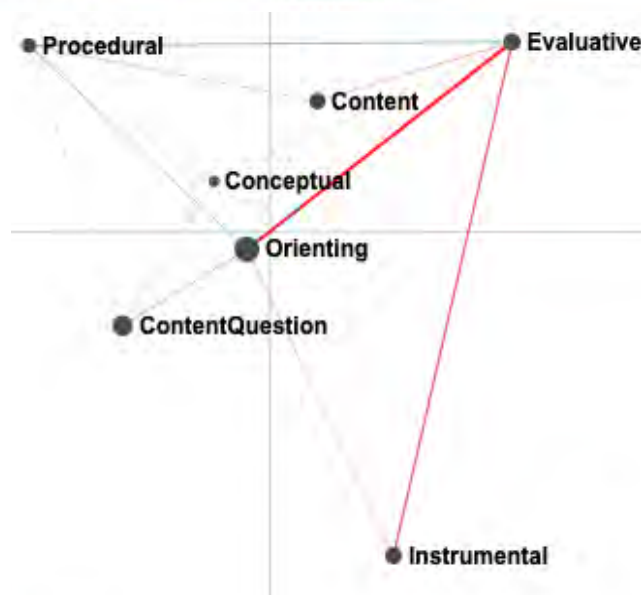


Figure 7. The subtracted network comparing when students improved (blue) or not (red) from the first to the third opportunity.

These findings are intuitively plausible since lower-level supports may lead to initial improvement, but they are unlikely to result in robust effects. An orienting support is important for situating an upcoming comment from the teacher, but a follow-up high-level support is needed to further engage the student in critical thinking. For example, telling a student, “No that’s not right, remember you are supposed to run a controlled experiment, so only change the amount of salt you are adding in the water” may help the student to complete the current activity or a very similar activity, whereas a combination of inquiry scaffolds such as “Remember you are supposed to run a controlled experiment and controlled trials means only changing one variable at once, so which variable should you change?” can be more beneficial to students’ application of this practice in later, varying activities. It is important to note that while a majority of students improved overall across the two teachers and this improvement was robust, there are a number of differences between the settings of the teachers’ classrooms, including different middle schools, different locations in the United States, and different numbers of students in each classroom, that could influence how each teacher provided support to their students in response to the dashboard. We examine differences between how the two teachers supported students in the following analyses.

4.2. Epistemic Networks by Teacher

Finally, we explored whether there were differences in the patterns of support provided by Ms. A and Mr. B in response to alerts from Inq-Blotter. This is important to consider since a majority of students in each classroom improved after being helped by their teacher, but it is unclear whether this improvement is because the dashboard elicited the same types of supports from each teacher. It could be that each teacher responded differently to the dashboard, but their students still similarly improved after receiving the relative teacher support informed by the dashboard. We constructed an epistemic network for each teacher using the whole conversation as the unit of analysis and the seven discourse codes as the nodes. An independent-samples t-test revealed that the mean centroid value for Mr. B ($M = 0.19, SD = 0.21$) was significantly different from the mean centroid value for Ms. A ($M = -0.19, SD = 0.25$), $t(39) = 5.27, p < 0.01, d = 1.67$, indicating a significant difference in the average pattern of feedback provided by each teacher.

It is important to note that both teachers provided a large proportion of orienting supports to their students in combination with other inquiry scaffolds (as aligns with the prior findings regarding the types of supports associated with improvement), but other differences did emerge in terms of the specific types of inquiry scaffolds. Mr. B did not provide any conceptual scaffolds to students (left of Figure 8), whereas Ms. A provided conceptual scaffolds in combination with primarily orienting scaffolds (right of Figure 8). While Mr. B did not integrate conceptual scaffolds, he did still apply other key inquiry scaffolds, including higher-level procedural support in combination with orienting supports (though not to the extent that Ms. A provided procedural scaffolds). Additionally, the subtracted network revealed that while both teachers provided orienting scaffolds to students, these scaffolds were combined with either primarily content-related and instrumental help in Mr. B’s case or evaluative comments in Ms. A’s case (see Figure 9). It appears that based on the results from the prior ENAs, the combination

majority of students improved on the science inquiry practice on which they were helped by a teacher using Inq-Blotter and that this improvement was significant. Additionally, we demonstrated that this improvement is maintained across multiple activities and that students who received help in response to dashboard alerts improved at a greater rate than students who did not receive teacher support. These findings are a significant contribution to the state of the field regarding the development of science dashboards (e.g., Acosta & Slotta, 2018; Matuk et al., 2016; Tissenbaum & Slotta, 2019) by demonstrating the potential of contextual alerts that directly address inquiry practices. Specifically, our findings elucidate the pedagogical benefits of alerting based on fine-grained assessment data of students' inquiry competencies.

We also analyzed the teacher supports that were provided in response to alerts from Inq-Blotter in order to understand how teacher support was fostering student improvement. Specifically, we identified differences in the patterns of support related to student improvement using ENA, which allowed for examining dynamic patterns in the discourse. The results showed that while lower-level scaffolds (orienting scaffolds) may benefit students' short-term inquiry performance, higher-level inquiry scaffolds (procedural and conceptual scaffolds) combined with content help were found to be most robust. This could be due to the value of using an orienting comment to situate a corresponding conceptual/procedural comment that engages a student in higher-level thinking related to the inquiry practice at hand. We also used ENA to compare the support patterns of the two participating teachers in order to understand how or whether individual teachers' supports differed while still benefiting student learning. Here we found that teacher support patterns did differ, suggesting that certain combinations of supports may function differently based on particular classroom settings but that in general the dashboard is effective in guiding these context-specific supports. The use of this in-depth analysis builds on prior innovative dashboard evaluation studies (e.g., Holstein et al., 2018a, 2018b, 2019; Martinez-Maldonado et al., 2013; Molenaar & Knoop-van Campen, 2017; Tissenbaum & Slotta, 2019) through an examination of the discourse data associated with dashboard use as well as through the triangulation of this discourse with other forms of data (i.e., student log file data, dashboard log file data).

In terms of our current and future work, the results of this study are being used to inform the design of inquiry supports for teachers (Adair, Dickler, & Gobert, 2020) that are being integrated into Inq-Blotter to help guide teachers in providing finer-grained, high-level support on inquiry practices to foster robust student improvement. This is especially important given that it is challenging to support students on NGSS practices (Pruitt, 2014), though teacher support has the potential to have a significant impact on student learning, as was demonstrated herein. In future studies, we will be including a larger number of teacher participants and examining whether any specific support patterns emerge across teachers in order to further expand upon the finding that there were significant differences in teacher support. In particular, it will be important to attend to how differences in support may relate to contextual factors (e.g., school demographics, location, etc.) and to teachers' corresponding understandings of the needs of the students with consideration of students' backgrounds. Also, with a larger number of participants, we will be able to explore differences in the patterns of support for each individual inquiry practice (hypothesizing, collecting data, and analyzing and interpreting data) and for particular virtual lab topics (flower growth, lunar phases, sound waves, etc.). This is important in order to account for potential differences in the types of support provided on each practice as well as differences in types of support provided for topics of varying complexity.

There are limitations to be addressed in future work, including that causation cannot be assumed entirely from the analyses in this work. For example, it is possible that teacher help that was not captured through log files or voice recording could, in part, explain students' learning gains, though this is unlikely because teachers used the technology to identify which students needed support. Teachers may also have provided certain supports to students based on their background understanding of their students' difficulties in addition to the information provided within the technology. In the future, we will also explore how teachers adjust their pedagogical practices for whole-class instruction in response to an alert. Additionally, we will explore the robustness of teacher support, elicited by alerts, on students' future performance on inquiry across longer time periods (e.g., several months) to also capture the influence of teachers supporting the same students multiple times and potentially on different practices.

Overall, this study illustrates *how* an alerting dashboard can support teachers and the corresponding influence on student learning. The findings from these analyses have the potential to inform future iterations of the technology as well as to examine the fidelity of the technology when implemented in classrooms with students. In conclusion, this study shows how an alerting dashboard could inform teacher support of students' science inquiry practice competencies and the value of in-depth analyses of the implementation of new technological genres in classrooms.

Declaration of Conflicting Interest

The first author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. The second and third authors are co-founders of the company that is commercializing Inq-ITS and Inq-Blotter.

Funding

This material is based upon work supported by the National Science Foundation under grant numbers 1629045 and 1902647, Institute of Education Sciences grant numbers EDIES16C0014 and EDIES15C0018, and a P.E.O. Scholar Award.

Acknowledgements

We greatly appreciated the thoughtful feedback and suggestions from the anonymous reviewers on this paper.

References

- Acosta, A., & Slotta, J. D. (2018). CKBiology: An active learning curriculum design for secondary biology. *Frontiers in Education*, 3, 1–19. <https://doi.org/10.3389/educ.2018.00052>
- Adair, A., Dickler, R., & Gobert, J. (2020). Supporting teachers supporting students: Evidence-based TIPS in a dashboard to guide inquiry scaffolding. In *Proceedings of the 14th International Conference of the Learning Sciences (ICLS 2020)*, 19–23 June 2020, Nashville, TN, USA (pp. 1769–1770). ISLS. Retrieved from <https://repository.isls.org/bitstream/1/6434/1/1769-1770.pdf>
- Aleven, V., Roll, I., McLaren, B. M., & Koedinger, K. R. (2016). Help helps, but only so much: Research on help seeking with intelligent tutoring systems. *International Journal of Artificial Intelligence in Education*, 26(1), 205–223. <https://doi.org/10.1007/s40593-015-0089-1>
- Charleer, S., Klerkx, J., & Duval, E. (2014). Learning dashboards. *Journal of Learning Analytics*, 1(3), 199–202. <https://doi.org/10.18608/jla.2014.13.22>
- Dickler, R., Gobert, J., & Sao Pedro, M. (2019a). Using epistemic network analysis to characterize teacher discourse in response to an alerting dashboard. Paper presented at the 29th Annual Meeting of the Society for Text and Discourse (ST&D 2019), 9–11 July 2019, New York.
- Dickler, R., Li, H., & Gobert, J. (2019b). Teacher scaffolds mediated by a science inquiry dashboard. Paper presented at the 14th Conference of the European Science Education Research Association (ESERA 2019), 26–30 August 2019, Bologna, Italy.
- Dillenbourg, P. (2013). Design for classroom orchestration. *Computers & Education*, 69, 485–492. <https://doi.org/10.1016/j.compedu.2013.04.013>
- Gee, J. P. (2004). *An Introduction to discourse analysis: Theory and method*. London, UK: Routledge. <https://doi.org/10.4324/9780203005675>
- Gobert, J. D., Baker, R. S., & Sao Pedro, M. A. (2014). *Inquiry skills tutoring system*. U.S. Patent No. 9,373,082. Washington, DC: U.S. Patent and Trademark Office.
- Gobert, J., Moussavi, R., Li, H., Sao Pedro, M., & Dickler, R. (2018). Real-time scaffolding of students' online data interpretation during inquiry with Inq-ITS using educational data mining. In A. K. M. Azad, M. Auer, A. Edwards, & T. de Jong (Eds.), *Cyber-physical laboratories in engineering and science education* (pp. 191–217). Cham: Springer. https://doi.org/10.1007/978-3-319-76935-6_8
- Gobert, J. D., & Sao Pedro, M. A. (2017). Digital assessment environments for scientific inquiry practices. In A. A. Rupp & J. P. Leighton (Eds.), *The Wiley handbook of cognition and assessment: Frameworks, methodologies, and applications* (pp. 508–534). West Sussex, UK: Wiley. <https://doi.org/10.1002/9781118956588.ch21>
- Gobert, J., Sao Pedro, M., Betts, C., & Baker, R. S. (2016). *Inquiry skills tutoring system (child patent for alerting system)*. U.S. Patent No. 9,564,057. Washington, DC: U.S. Patent and Trademark Office.
- Gobert, J. D., Sao Pedro, M., Raziuddin, J., & Baker, R. S. (2013). From log files to assessment metrics: Measuring students' science inquiry skills using educational data mining. *Journal of the Learning Sciences*, 22(4), 521–563. <https://doi.org/10.1080/10508406.2013.837391>
- Holstein, K., McLaren, B. M., & Aleven, V. (2018a). Student learning benefits of a mixed-reality teacher awareness tool in AI-enhanced classrooms. In *Proceedings of the 19th International Conference on Artificial Intelligence in Education (AIED 2018)*, 27–30 June 2018, London, UK (pp. 154–168). Cham: Springer. https://doi.org/10.1007/978-3-319-93843-1_12
- Holstein, K., McLaren, B. M., & Aleven, V. (2018b). Informing the design of teacher awareness tools through causal alignment analysis. In *Proceedings of the 13th International Conference of the Learning Sciences (ICLS 2018)*, 23–27 June 2018, London, UK (pp. 104–111).
- Holstein, K., McLaren, B. M., & Aleven, V. (2019). Co-designing a real-time classroom orchestration tool to support teacher–AI complementarity. *Journal of Learning Analytics*, 6(2), 27–52. <https://doi.org/10.18608/jla.2019.62.3>

- Howe, C., Hennessy, S., Mercer, N., Vrikki, M., & Wheatley, L. (2019). Teacher–student dialogue during classroom teaching: Does it really impact on student outcomes? *Journal of the Learning Sciences*, 28(4-5), 462–512. <https://doi.org/10.1080/10508406.2019.1573730>
- Jurow, A. S., & Creighton, L. (2005). Improvisational science discourse: Teaching science in two K-1 classrooms. *Linguistics and Education*, 16(3), 275–297. <https://doi.org/10.1016/j.linged.2006.02.002>
- Laurillard, D. (2008). The teacher as action researcher: Using technology to capture pedagogic form. *Studies in Higher Education*, 33(2), 139–154. <https://doi.org/10.1080/03075070801915908>
- Manz, E., & Renga, I. P. (2017). Understanding how teachers guide evidence construction conversations. *Science Education*, 101(4), 584–615. <https://doi.org/10.1002/sce.21282>
- Marquart, C. L., Hinojosa, C., Swiecki, Z., Eagan, B., & Shaffer, D. W. (2018). *Epistemic network analysis (version 1.5.2)*. Retrieved from <https://www.epistemicnetwork.org/>
- Martinez-Maldonado, R., Clayphan, A., Yacef, K., & Kay, J. (2015). MTFeedback: Providing notifications to enhance teacher awareness of small group work in the classroom. *IEEE Transactions on Learning Technologies*, 8(2), 187–200. <https://doi.org/10.1109/TLT.2014.2365027>
- Martinez-Maldonado, R., Kay, J., Yacef, K., Edbauer, M. T., & Dimitriadis, Y. (2013). MTClassroom and MTDashboard: Supporting analysis of teacher attention in an orchestrated multi-tabletop classroom. In *Proceedings of the 10th International Conference on Computer Supported Collaborative Learning (CSCL 2013)*, 15–19 June 2013, Madison, WI, USA. (pp. 320–327).
- Martinez-Maldonado, R., Kay, J., Yacef, K., & Schwendimann, B. (2012). An interactive teacher’s dashboard for monitoring groups in a multi-tabletop learning environment. In *Proceedings of the International Conference on Intelligent Tutoring Systems (ITS 2012)*, 14–18 June 2012, Chania, Greece (pp. 482–492). Springer. https://doi.org/10.1007/978-3-642-30950-2_62
- Matuk, C., Gerard, L., Lim-Breitbart, J., & Linn, M. (2016). Gathering requirements for teacher tools: Strategies for empowering teachers through co-design. *Journal of Science Teacher Education*, 27(1), 79–110. <https://doi.org/10.1007/s10972-016-9459-2>
- Mislevy, R. J., Behrens, J. T., Dicerbo, K. E., & Levy, R. (2012). Design and discovery in educational assessment: Evidence-centered design, psychometrics, and educational data mining. *Journal of Educational Data Mining*, 4(1), 11–48. <https://doi.org/10.5281/zenodo.3554641>
- Molenaar, I., & Knoop-van Campen, C. (2017). Teacher dashboards in practice: Usage and impact. In *Proceedings of the 12th European Conference on Technology Enhanced Learning (EC-TEL 2017)*, 12–15 September 2017, Tallinn, Estonia (pp. 125–138). Springer, Cham. https://doi.org/10.1007/978-3-319-66610-5_10
- Molenaar, I., & Knoop-van Campen, C. (2018). Experience matters: The impact of dashboards on teachers’ feedback. In *Proceedings of the 13th International Conference of the Learning Sciences (ICLS ’18)*, 23–27 June 2018, London, UK (pp. 1231–1232).
- Next Generation Science Standards Lead States. (2013). *Next generation science standards: For states, by states*. The National Academies Press. <https://doi.org/10.17226/18290>
- Pruitt, S. L. (2014). The next generation science standards: The features and challenges. *Journal of Science Teacher Education*, 25(2), 145–156. <https://doi.org/10.1007/s10972-014-9385-0>
- Roschelle, J., Martin, W., Ahn, J., & Schank, P. (Eds.) (2017). *Cyberlearning community report: The state of cyberlearning and the future of learning with technology* (technical report). SRI International. Retrieved from <http://circlcenter.org/wp-content/uploads/2017/07/CyberlearningCommunityReport2017.pdf>
- Rupp, A. A., Choi, Y., Gushta, M., Mislevy, R., Thies, M.C., Bagley, E., ... Shaffer, D. (2009). Modeling learning progressions in epistemic games with epistemic network analysis: Principles for data analysis and generation. In *Proceedings of the 2009 Learning Progressions in Science Conference (LeaPS 2009)*, 24–26 June 2009, Iowa City, IA, USA (pp. 24–26). Retrieved from <https://www.education.msu.edu/projects/leaps/proceedings/Rupp.pdf>
- Sao Pedro, M. A., Gobert, J., & Dickler, R. (2019). Can an alerting teacher dashboard improve how teachers help their students learn science inquiry practices? Paper presented at the 2019 American Educational Research Association Annual Meeting: Learning and Instruction (AERA 2019), 5–9 April 2019, Toronto, ON, Canada.
- Schwarz, B. B., Prusak, N., Swidan, O., Livny, A., Gal, K., & Segal, A. (2018). Orchestrating the emergence of conceptual learning: A case study in a geometry class. *International Journal of Computer-Supported Collaborative Learning*, 13(2), 189–211. <https://doi.org/10.1007/s11412-018-9276-z>

- Segal, A., Hindi, S., Prusak, N., Swidan, O., Livni, A., Palatnic, A., & Schwarz, B. (2017). Keeping the teacher in the loop: Technologies for monitoring group learning in real-time. In *Proceedings of the 18th International Conference on Artificial Intelligence in Education (AIED 2017)*, 28 June–1 July 2017, Wuhan, China (pp. 64–76). Cham: Springer. https://doi.org/10.1007/978-3-319-61425-0_6
- Shaffer, D. W., Collier, W., & Ruis, A. R. (2016). A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *Journal of Learning Analytics*, 3(3), 9–45. <https://doi.org/10.18608/jla.2016.33.3>
- Sherin, M., Jacobs, V., & Philipp, R. (Eds.). (2011). *Mathematics teacher noticing: Seeing through teachers' eyes*. London, UK: Routledge.
- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 7, 1–22. <https://doi.org/10.17763/haer.57.1.j463w79r56455411>
- Svarovsky, G. N. (2011). Exploring complex engineering learning over time with epistemic network analysis. *Journal of Pre-College Engineering Education*, 1(2), Article 4. <https://doi.org/10.5703/1288284314638>
- Tissenbaum, M., & Slotta, J. (2019). Supporting classroom orchestration with real-time feedback: A role for teacher dashboards and real-time agents. *International Journal of Computer-Supported Collaborative Learning*, 14(3), 325–351. <https://doi.org/10.1007/s11412-019-09306-1>
- Tabak, I., & Kyza, E. A. (2018). Research on scaffolding in the learning sciences: A methodological perspective. In F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, & Peter Reimann (Eds.), *International handbook of the learning sciences* (pp. 191–200). Routledge. <https://doi.org/10.4324/9781315617572-19>
- van Leeuwen, A. & Rummel, N. (2018). Considering teachers' informational needs and the role division between teacher and orchestration tool in the context of collaborative learning. In *Proceedings of the 13th International Conference of the Learning Sciences (ICLS 2018)*, 23–27 June 2018, London, UK (pp. 1230–1231).
- van Leeuwen, A., van Wermeskerken, M., Erkens, G., & Rummel, N. (2017). Measuring teacher sense making strategies of learning analytics: A case study. *Learning: Research and Practice*, 3(1), 42–58. <https://doi.org/10.1080/23735082.2017.1284252>
- 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>
- West, D. M. (2012). *Big data for education: Data mining, data analytics, and web dashboards*. Governance Studies at Brookings. Retrieved from <https://www.brookings.edu/wp-content/uploads/2016/06/04-education-technology-west.pdf>