Analyzing Student-Teacher Discourse Prompted by a Real-Time Alerting Dashboard for Science Inquiry Practices

Objective

The present study used an analysis of teacher-student discourse to examine how conversations prompted by fine-grained alerts on science inquiry practices in the state-of-the-art teacher dashboard, [DASHBOARD], related to changes in student inquiry performance within virtual labs in the intelligent tutoring system, [ITS].

Theoretical Framework

The Next Generation Science Standards (NGSS, 2013) emphasize science inquiry, but it remains a challenge for teachers to assess and support students' competencies on the practices (McElhaney et al., 2017; Pruitt, 2014). Developments in educational technology, such as virtual science labs with automated scoring of student performance (e.g., Quellmalz et al., 2012; Vitale et al., 2016), provide an opportunity to assess students in real-time and provide teachers with the information needed to support students' competencies.

Researchers have developed teacher dashboard technologies (i.e., tools that report on student progress within an online activity; Dillenbourg et al., 2013) to accompany virtual science activities (e.g., Authors, 2018; Matuk et al., 2016; Tissenbaum & Slotta, 2019). This allows teachers to work closely with students in the role of a facilitator versus lecturer (Holstein et al., 2019). Existing dashboards that have been developed for science classrooms, however, focus primarily on science content (e.g., Acosta & Slotta, 2018; Matuk et al., 2016; Tissenbaum & Slotta, 2019). Here we argue that, in order to realize the vision of NGSS, teachers need access to fine-grained data on students' inquiry performance in real time so that they can better support them on the practices while students work. The teacher dashboard, [DASHBOARD], meets this need by sending real-time, actionable alerts to teachers regarding students' *science inquiry* competencies, which are assessed using educational data-mined and knowledge-engineered algorithms in the intelligent tutoring system [ITS] (Authors, 2018). It is important to examine the use and impact of this new technological innovation in classrooms (e.g., Laurillard, 2008) for science teaching and learning.

While researchers have explored discourse around dashboard use (i.e., Molenaar & Knoop Van-Campen, 2018; Tissenbaum & Slotta, 2019), few studies have examined how teacher-student discourse promotes students' inquiry learning. Prior studies in science classrooms have shown that coding discourse for both the dynamics of conversations (i.e., contributions of teachers relative to students; Hardman, 2016; Manz & Renga, 2017; Pimentel & McNeill, 2013) and the content (i.e., the type of information exchanged; Howe et al., 2019; Smart & Marshall, 2013) is illuminating in terms of understanding student learning outcomes (Wilson et al., 2011). Further, studies revealed that teachers can benefit from pedagogical supports that promote high-level, rich dialogue centered on practices (e.g., McNeill et al., 2016; Smart & Marshall, 2013). Dashboards for science inquiry, such as [DASHBOARD] (Authors, 2018), present an opportunity to build on these findings by providing teachers with the fine-grained, inquiry practice performance data that is needed to facilitate productive discourse that supports learning.

Prior randomized controlled studies with [DASHBOARD] demonstrated that students showed greater improvement when teachers used alerts, compared to classrooms where teachers did not have access to real-time alerts (Authors, 2018, 2019); however, studies have not explored students' contributions to discourse (which may elucidate the processes by which their knowledge of the practices is reified; e.g., Smart & Marshall, 2013). In the present study, we sought to examine the student-teacher discourse prompted by [DASHBOARD] alerts in relation to student learning with the following research questions (RQs):

RQ1: When teachers use [DASHBOARD] to help students on an inquiry practice, do students improve on that practice on their next opportunity?

RQ2: How do the discourse dimensions of teacher-student contributions and extent of inquiry practice discussion during discourse relate to changes in student inquiry performance?

Methods

Participants

Four science teachers from two American middle schools implemented [DASHBOARD] while their students used [ITS] during science class periods in 2018-2019. Twenty-five students were helped by a teacher based on a [DASHBOARD] alert and then continued on to complete a second activity in [ITS].

Materials

Students completed virtual physical science labs in [ITS] (Authors, 2013) that were aligned to the practices outlined in the NGSS Framework (2013). Specifically, each [ITS] lab consisted of four stages: asking questions/hypothesizing, carrying out investigations/collecting data, analyzing and interpreting data, and communicating findings (Figure 1).

Teachers used [DASHBOARD] (Authors, 2018) on a tablet and received real-time alerts on students' difficulties with the practices in [ITS] at a fine-grained, actionable level (Figure 2). Alerts for teachers in [DASHBOARD] are triggered by automated scoring of student performance within the first three stages of [ITS] (alerts for the final stage are in development; Authors, 2017) using algorithms validated in prior studies (see measures section; Authors, 2013, 2018). Dashboard alerts include information about both the practice and the activity on which the student is working. For example, if a student is having difficulty analyzing data (Figure 2, center panel), the teacher can click on the alert and see a detailed breakdown (Figure 2, right panel) of how the student has not selected sufficient evidence to support their claim (a common student difficulty; i.e., Sandoval & Millwood, 2005).

Procedure

After being provided with an overview of [DASHBOARD] and [ITS], teachers introduced [ITS] to their students and employed [DASHBOARD] while students completed at least three virtual lab investigations. The teachers were instructed to record voice data any time that they helped a student based on an alert (Figure 2, right panel).

Measures

All students' actions in [ITS] were automatically scored based on previously validated educational data-mined and knowledge-engineered algorithms (Authors, 2013, 2016, 2018). The practices were scored based on multiple components (the "sub-practices") aligned to NGSS, and refined through iterative development and testing in prior studies (Table 1; Authors, 2012, 2013, 2016, 2018). Inquiry practice scores, along with timestamps, were logged.

As teachers used [DASHBOARD], the system captured and stored data such as the content of the alerts, the student ID, the voice data of the interaction, and the corresponding timestamps. This data was essential for the analyses in terms of identifying which students were helped by the teachers and the practice on which they were helped.

Analyses

We triangulated the student and teacher log files, based on timestamps, to identify all instances in which: a teacher helped a student based on an alert in [DASHBOARD], a recording of the discourse was captured, and the student went on to complete a second virtual lab activity (N = 25). To answer RQ1, we examined the inquiry practice scores of the 25 students prior to being helped by the teacher and their scores in the next virtual lab activity after being helped. To answer RQ2, the voice recordings of teacher-student conversations were transcribed and a coding scheme was developed based on science classroom discourse literature (Howe et al., 2019; Pimentel & McNeill, 2013; Smart & Marshall, 2013). The final coding scheme had two scaled dimensions (see Table 2 for further examples):

- Dynamics - the level of teacher and student contributions: (1) teacher-led, (2) equally led by teacher and student, (3) student-led.
- Practices - the extent of inquiry practice discussion: (1) no inquiry practice discussion, (2) relevant practice discussion, (3) practices and sub-practices were *explicitly* discussed/mentioned.

Each conversation was assigned one code for each dimension. Five researchers coded the first six conversations together. Two researchers then independently scored another ten conversations with 70% agreement. Discrepancies were addressed and the remaining conversations were scored with 100% agreement.

Employing the triangulated data and coded conversations, we conducted a linear regression analysis (e.g., Smart & Marshall, 2013) to examine if the two discourse dimensions could reliably predict change in inquiry practice performance. We then explored specific examples of discourse that exemplified the patterns found in the quantitative analyses.

Results

To answer RQ1, we first calculated the change in performance score (i.e., performance on a practice prior to being helped subtracted from performance on the practice in the next activity after being helped) for each student. We found a positive change in performance for 80% of students (21 of 25 students improved on the inquiry practice after being helped by the teacher in response to [DASHBOARD]; see Table 1). For 30% of students, there was no change in performance (2 students) or the change in performance was negative (2 students). In order to further investigate these findings, it was essential to analyze the discourse that occurred when the teachers responded to the alerts.

To answer RQ2, we conducted a regression analysis to examine the relationship between the dimensions of student-teacher discourse (see Table 2) and change in performance. The two discourse dimensions (dynamics and practices scaled from 1 to 3) represented the predictor variables and the change in performance represented the dependent variable. Results indicated that the linear regression model was marginally significant and could explain 23% of the variance in the data, $F(2,23) = 3.41$, $p = .050$, $R^2 = .23$ (based on $\alpha \leq .05$). While both predictor variables were positively correlated with an increase in performance, the only variable that was a significant predictor of change in performance was dynamics, $t(25) = 2.36$, $p = .027$, $\theta = .32$. This finding indicates that conversations with more student contributions were associated with greater improvement in student inquiry performance. Discourse focused on practices was not a significant predictor, $t(25) = 1.01$, $p = .323$, $\theta = .13$, but was positively correlated with change in performance, indicating that more explicit practice discussion was related to greater improvement in inquiry performance. Overall these results indicate that higher levels of both teacher-student discourse and practice discussions are associated with student improvement on inquiry practice competencies. These findings align with the science inquiry discourse literature (e.g., McNeill et al., 2016; Smart & Marshall, 2013), which identified the relevance of students' contributions and levels of inquiry discussion to higher learning outcomes.

To further examine the data in our quantitative analyses, we analyzed conversations that illustrated the contribution of the different discourse dimensions to student inquiry learning. For example, the following excerpt occurred in response to a [DASHBOARD] alert that a student was struggling to select controlled trials as evidence for their claim:

> **Student:** It's not letting me select my evidence.
>
> **Teacher:** Okay, so what did you, what are you doing?
>
> **Student:** If I change the length of the string it will change the loudness.
>
> **Teacher:** Okay, let's look at your data …Everything should be the same except the length of the string, right?
>
> **Student:** Yeah…
>
> **Teacher:** So only this [Length of the String] column should change.

This example depicts discourse with meaningful teacher and student contributions that also explicitly discuss sub-practices related to analyzing data ("select my evidence" , "look at your data"; see Table 1). The teacher reminds the student to only select data that is controlled (i.e., only the Length of the String column changing) as evidence to support their claim. The students' new understanding is reflected in their improved performance from prior to being helped (44% on analyzing data) to their next activity (100% on analyzing data). Additional examples of teacher-student discourse were identified for the practices of hypothesizing and collecting data (Table 3). These examples characterize and illustrate the conclusion that dynamic conversations containing explicit practice/sub-practice discussion in response to [DASHBOARD] alerts could predict improvements in inquiry practice performance.

Scholarly Significance

We found that a model of discourse characterized by greater student contributions and explicit practice discussion prompted by an innovative science inquiry technology could significantly predict improvement in student competencies on subsequent inquiry tasks. It is valuable to understand how teacher-student interactions facilitated by technology relate to student outcomes in order to inform design iterations (i.e., future [DASHBOARD] designs will embed additional support within alerts to prompt richer practice discussions; Authors, 2020). The present study demonstrates how technologies like [ITS] and [DASHBOARD] can support teachers' pedagogical practices and students' inquiry learning. The fine-grained alerts on students' competencies provide teachers with the actionable data needed to address student difficulties, as evidenced by students' improvement on the practices on which they were helped. This work demonstrates the value of student-teacher discourse analyses in un-packing the impact of implementing innovative technologies for science teaching and learning.

References

Acosta, A., & Slotta, J. D. (2018). CKBiology: An active learning curriculum design for secondary biology. *Frontiers in Education, 3*, 1-19.

Authors. (2012, 2013, 2016, 2017, 2018, 2019, 2020).

Dillenbourg, P. (2013). Design for classroom orchestration. *Computers & Education, 69*, 485-492.

Hardman, J. (2016). Opening-up classroom discourse to promote and enhance active, collaborative and cognitively-engaging student learning experiences. In C. Goria, O. Speicher, & S. Stollhans (Eds.)*, Innovative language teaching and learning at university: enhancing participation and collaboration* (pp. 5-16). Dublin: Research-Publishing.net.

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.

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.

Laurillard, D. (2008). The teacher as action researcher: using technology to capture pedagogic form. *Studies in Higher Education*, *33*(2), 139-154.

Manz, E., & Renga, I. P. (2017). Understanding how teachers guide evidence construction conversations. *Science Education*, *101*(4), 584-615.

Matuk, C. F., Linn, M. C., & Eylon, B. S. (2016). Technology to support teachers using evidence from student work to customize technology-enhanced inquiry units. *Instructional Science*, *43*(2), 229-257.

McElhaney, K. W., Gane, B. D., DiBello, L. V., Fujii, R., Pennock, P. H., Vaishampayan, G., & Pellegrino, J. W. (2017). Designing scoring rubrics to support NGSS-aligned, classroom-based formative assessment. In, *Proceedings of the 2017 American Educational Research Association Annual Meeting*, San Antonio, TX.

McNeill, K. L., González-Howard, M., Katsh-Singer, R., & Loper, S. (2016). Pedagogical content knowledge of argumentation: Using classroom contexts to assess high-quality PCK rather than pseudoargumentation. *Journal of Research in Science Teaching*, *53*(2), 261-290.

Molenaar, I., &. Knoop-van Campen, C. (2018). Experience matters: The impact of dashboards on teachers' feedback. In, *Proceedings of the 2018 International Conference of the Learning Sciences* (pp. 1231-1232).

NGSS Lead States. (2013). *Next Generation Science Standards: For States, By States.* Washington, DC: The National Academies Press.

Pimentel, D. S., & McNeill, K. L. (2013). Conducting talk in secondary science classrooms: Investigating instructional moves and teachers' beliefs. *Science Education*, *97*(3), 367-394.

Pruitt, S. L. (2014). The next generation science standards: The features and challenges. *Journal of Science Teacher Education, 25*(2), 145-156.

Quellmalz, E. S., Timms, M. J., Silberglitt, M. D., & Buckley, B. C. (2012). Science assessments for all: Integrating science simulations into balanced state science assessment systems. *Journal of Research in Science Teaching, 49*(3), 363-393.

Sandoval, W. A., & Millwood, K. A. (2005). The quality of students' use of evidence in written scientific explanations. *Cognition and Instruction*, *23*(1), 23-55.

Smart, J. B., & Marshall, J. C. (2013). Interactions between classroom discourse, teacher questioning, and student cognitive engagement in middle school science. *Journal of Science Teacher Education, 24*(2), 249-267.

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.

Vitale, J., Linn, M., & Gerard, L. (2016). Visualizing data from automated scores to help teachers guide inquiry with scientific visualizations in diverse classes. In proceedings of the *12th Annual International Conference of the Learning Sciences*, Singapore.

Wilson, K., Collins, G., Couchman, J., & Li, L. (2011). Co-constructing academic literacy: Examining teacher-student discourse in a one-to-one consultation. *Journal of Academic Language and Learning*, *5*(1), A139-A153.

Figures and Tables

Figure 1. Screenshots the stages of an [ITS] virtual lab

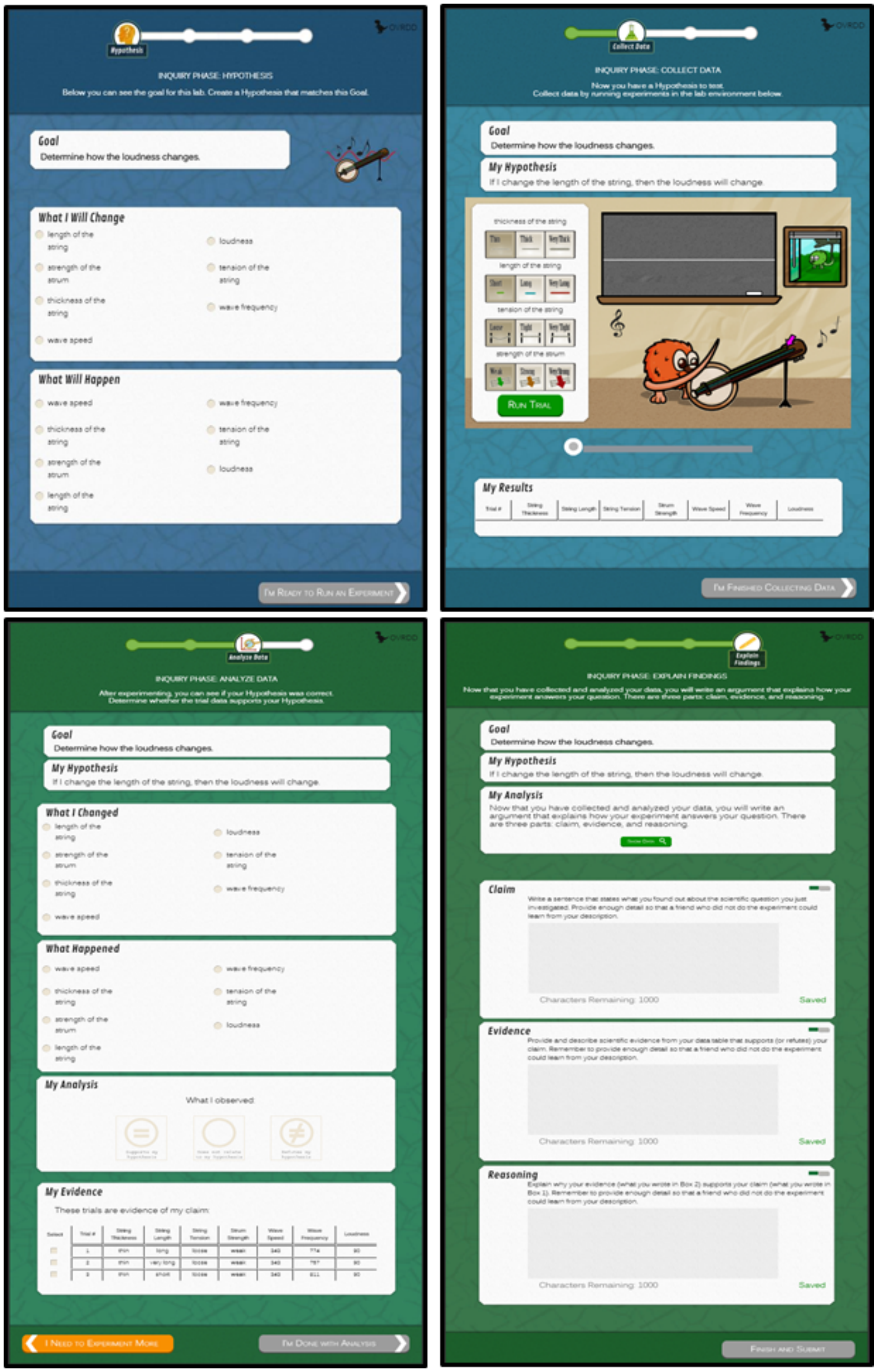

Figure 2. Screenshot of the [DASHBOARD] system with a selected alert

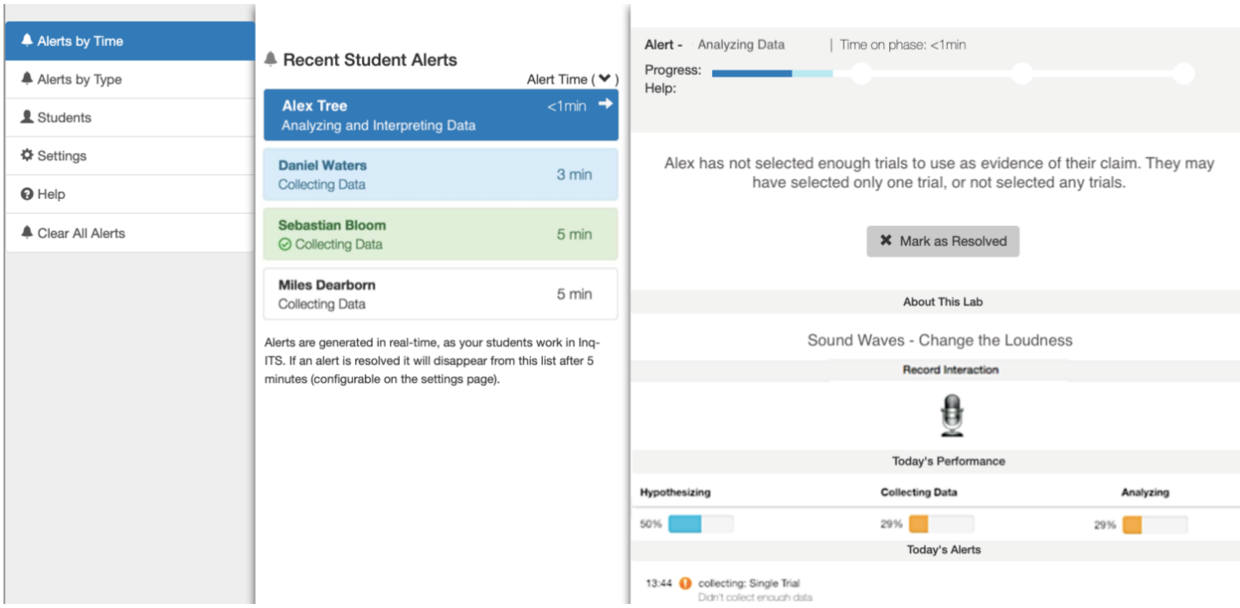

Table 1. Practices scored and alerted on in the first three stages of [ITS]

Inquiry Practice	Examples of Inquiry Sub-Practices	# of Students Helped on Practice	Average Change in Performance
Asking Questions/ Hypothesizing	Identifying independent and dependent variables to investigate; Selecting variables appropriate for the investigation goal	8	+ 18%
Carrying out Investigations/ Collecting Data	Running sufficient and controlled trials; Investigating the target independent variable	5	+ 36%
Analyzing and Interpreting Data	Making an accurate claim about the relationship between the independent and dependent variables based on the investigation; Selecting sufficient and controlled trials related to the investigation as evidence to support the claim	12	+ 24%

Table 2. Discourse coding scheme with two dimensions

Dimension	Level	Codes	Description	Frequency (% of Data)	Example Conversation Excerpt (T: Teacher, S: Student)
Dynamics	1	Teacher-Led	The teacher provides the majority of the contributions in the discourse.	12 (48%)	**T:** Okay, alright. In your experiment changing the strength of the strum did not cause waves...okay. So go back up to your claim. Now go down a little bit. Okay, you said the wave speed decreased, did it? **S:** Ohhh. **T:** Do you, does that make sense now?
	2	Teacher-Student	The teacher and student make relatively equal contributions to the discourse.	13 (52%)	**T:** Measure the loudness? **S:** Yup. Maybe, the length of the string? **T:** Length of the string? Okay. Are you going to increase or decrease? **S:** Increase? Aren't we changing this?
	3	Student-Led	The student provides the majority of contributions to the discourse.	0 (0%)	None
Practices	1	No practice	No references made to inquiry practices, either directly or indirectly.	1 (4%)	[Alert in [DASHBOARD] that student is struggling to make a claim on relationship between variables] **S:** They create [the labs] and then... **T:** I mean I can look at them once they answer it and then I just kind of...
	2	Relevant Practice Discussion	Discourse is relevant to the practice alert received by the teacher, but there are no explicit mentions of sub-practices.	6 (24%)	[Alert in [DASHBOARD] that student is struggling to make a claim on relationship between variables] **T:** So determine how the wave speed changes. So do you think how long the string is would effect it? So that's saying, ya know, if you have a shorter string the wave speed will be faster…
	3	Explicit Practice/Sub-Practice Discussion	Discourse contains explicit references to practices *and* sub-practices mentioned in alert.	18 (72%)	[Alert in [DASHBOARD] that student is struggling to form hypothesis based on activity goal] **T:** Well...You're gonna determine how [the variable] changes right? **S:** Yes, so… **T:** Yeah, yeah....cause that's what you're measuring...so that one should be loudness. Let's look. So your gonna change the speed of the wave and that's gonna effect the loudness?

Table 3. Additional examples of teacher-student discourse

[DASHBOARD] Alert	Teacher (T) - Student (S) Discourse	Practice and Sub-Practice Comments	Student Performance Prior to help (Activity 1)	Student Performance After help (Activity 2)
Asking Questions/ Hypothesizing: Student selected variables in their hypothesis that are not testable	**T:** Are you having trouble with your hypothesis? **S:** Yes. Yeah. **T:** What are you doing? **S:** I tried to do the wave frequency one but that wasn't working. **T:** Okay, let me see your goal. **S:** Okay, the goal was to… **T:** Okay, so what are you gonna…okay, Determine how the wave speed changes. So you what…so you said the frequency? So now you're measuring…wave speed...	"hypothesis" "let me see your goal" "now you're measuring"	67% on Hypothesizing	100% on Hypothesizing
Carrying out Investigations/ Collecting Data: Student is struggling to run an experiment with sufficient trials that test their hypothesis	**S:** I chose the green ball because it had the highest mass and that would probably I think come out as better to work with the final velocity thing … **T:** But don't you think that you should probably try more trials? Because it's hard to analyze data when… **S:** There's more, there's more like these, right? **T:** Yes. **S:** OK, so I need to try more in the next one.	"try more trials" "analyze data"	56% on Collecting Data	100% on Collecting Data
Analyzing and Interpreting Data: Student is struggling to select data that aligns with their claim	**S:** It's not letting me select my evidence. **T:** Okay, so what did you, what are you doing **S:** If I change the length of the string it will change the loudness. **T:** Okay, let's look at your data …Everything should be the same except the length of the string, right? **S:** Yeah **T:** So only this column should change	"select my evidence" "look at your data"	44% on Analyzing and Interpreting Data	100% on Analyzing and Interpreting Data