# Multimodal Data Collection and Analysis of Collaborative Learning through an Intelligent Tutoring System

Ran Liu and John Stamper

Carnegie Mellon University, Pittsburgh PA 15201, USA
ranliu@cmu.edu, jstamper@cs.cmu.edu

**Abstract.** A great deal of learning analytics research has focused on what can be achieved by analyzing log data, which can yield important insights about how students learn in online systems. Log data cannot capture all important learning phenomena, especially in open-ended, collaborative, or project-based environments. Collecting and processing/analyzing additional multimodal data streams, however, present many methodological challenges. We describe two datasets from similar collaborative-learning oriented educational technologies deployed in classrooms but with different streams of multimodal data collected. We discuss the differing insights that have resulted from each study, due largely to the specific streams of multimodal data collected. We review the challenges that remain. Finally, we present methods we've developed to streamline the temporal alignment and linkage across multiple data streams.

**Keywords:** Intelligent Tutoring System, Collaborative Learning, Usage Logs; Multimodal data, Multimodal analytics

## 1 Introduction

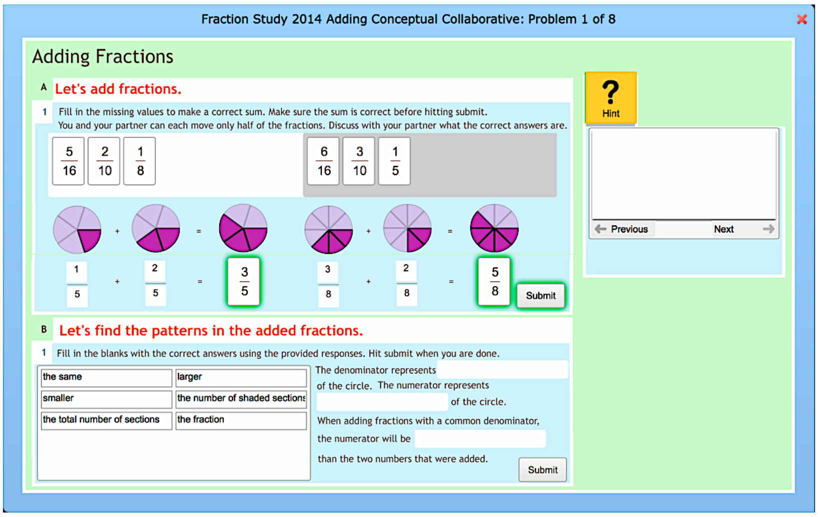
As education technology becomes more prevalent, large amounts of learning-related data are being produced. A great deal of learning analytics research has focused on what can be achieved by analyzing log data, which can yield important insights about how students learn in online systems. But log data cannot capture all important learning phenomena, especially those that take place in open-ended, collaborative, or project-based environments [1, 6]. Multimodal data streams that richly capture the context surrounding educational technology use may add to and complement log data. In some cases, they may lead to critical insights.

Learning analytics conducted on log data often omit additional contextual data for a number of reasons. Data on classroom context are difficult to collect. Data from different sources are often collected at different grain sizes, which are difficult to integrate. We present two datasets from similar educational technologies deployed in collaborative learning contexts but with different streams of multimodal data. In one study, we collected high-quality audio recordings of individual students as they engaged in collaborative dialogue, full-classroom video, and close-up focal video of two dyads. In the other study, we collected audio and screen video recordings of each student working on the tutor using Camtasia. We discuss the differing insights that have resulted from each study, due largely to the specific streams of multimodal data collected.

Additionally, we present methods we developed to streamline the synchronization and analysis of multimodal data streams. These open-source tools support the temporal alignment of software-logged usage data to multimodal data streams, visualization and exploratory analyses of aligned streams, and event-based extraction of video segments.

## 2 The Datasets

Both datasets were collected in classroom studies of students working on the Collaborative Fraction Tutor [5], an intelligent tutoring system developed by researchers at Carnegie Mellon University that helps students become better at understanding and working fractions. The tutor was created using Cognitive Tutor Authoring Tools, which facilitate rapid development and easy deployment of intelligent tutors. The tutor supports collaboration between partners in order to learn fraction skills such as addition (Figure 1), subtraction, comparing fractions to determine which is larger or smaller, finding the least common denominator, and finding equivalent fractions. Each student in a pair can control only part of the screen, so both partners must work together in order to finish the problem. Students work at the same time and can talk about what they are doing, ask for help from their partner, and generally collaborate to get the correct answer.



Fraction Study 2014 Adding Conceptual Collaborative: Problem 1 of 8

Adding Fractions

A Let's add fractions.

1 Fill in the missing values to make a correct sum. Make sure the sum is correct before hitting submit.
You and your partner can each move only half of the fractions. Discuss with your partner what the correct answers are.

$\frac{5}{16}$ $\frac{2}{10}$ $\frac{1}{8}$ $\frac{6}{16}$ $\frac{3}{10}$ $\frac{1}{5}$

$\frac{1}{5} + \frac{2}{5} = \frac{3}{5}$ $\frac{3}{8} + \frac{2}{8} = \frac{5}{8}$

B Let's find the patterns in the added fractions.

1 Fill in the blanks with the correct answers using the provided responses. Hit submit when you are done.

| | |
|---|---|
| the same | larger |
| smaller | the number of shaded sections |
| the total number of sections | the fraction |

The denominator represents of the circle. The numerator represents of the circle.
When adding fractions with a common denominator, the numerator will be than the two numbers that were added.

Fig. 1. Example screen from the Collaborative Fraction Tutor.

### 2.1 Dataset 1

**Collection.** Participants were 104 fourth and fifth graders from one middle and one elementary school in the greater Pittsburgh area. There were 19 fifth graders from the middle school, and 50 fourth graders and 35 fifth graders from the elementary school. Students participated across five 45-minute class periods on consecutive days within a week. On the first and last days, students took a computerized pre- and post-test, respectively. They engaged in the Collaborative Fraction Tutor during the three consecutive days between the pre- and post-test days.

Only a subset of 36 students (14 fifth graders from the middle school, and 16 fourth graders and 6 fifth graders from the elementary school) were present for the full study, had the same partner during the entire study (no absences for either individual), and consented to audio recordings of their dialogue. Obtaining opt-in consent for the relatively more invasive collection of multimodal data streams and the reduction in sample size for the full dataset are challenges we faced during data collection.

For the consenting students, high-quality audio data were collected for each individual student using a headset outfitted with a microphone. The microphone was linked to a tablet computer to store the recordings.

In each class, we also collected full-classroom video recorded from one camera located in the corner of the room. Finally, we collected two dyads worth of “focal” video (across all three days of tutor use, excluding the pretest and posttest) in which the video camera was positioned behind the dyad and pointed at the students’ computer screens.

**Analyses & Challenges Faced.** The main analyses we have done thus far with this dataset has been to professionally transcribe the audio data and conduct natural language processing analyses to relate the dialogue to learning outcomes (measured by pretest to posttest gains) [2].

One analysis challenge was that for these particular analyses, the transcripts needed to be at the dyad level. However, the recordings were collected at the individual level. Aligning and merging the recordings between the two individuals of each dyad required a significant amount of human effort.

We used the STREAMS tools we developed (described in the next section) to temporally align the focal students’ video files with the corresponding usage log files and dialogue transcripts. We also developed code to automatically import these time-synchronized data streams into DataVyu for easy visualization and additional coding.

One remaining challenge is that one temporal synchronization point must be manually inputted by a human for every data stream that must be synced. The amount of human effort required per data stream is minimal, but scales linearly with the quantity of data collected. Future methods that create automatic temporal synchronization points between different data streams during data collection would circumvent the need for this human time and effort.

### 2.2 Dataset 2

**Collection.** Participants were 26 fifth grade students at a middle school in the greater Pittsburgh area enrolled in an advanced math class. Students participated across five 45-minute class periods on consecutive days within a week. On the first and last days, students took a computerized pre- and post-test, respectively. They engaged in the Collaborative Fraction Tutor during the three consecutive days between the pre- and post-test days. Students spent half of each class period working individually and half collaborating with a partner. Students were paired with the same person for all partner activities throughout the experiment. We collected Camtasia screen video and audio captures for all students across all three days of tutor use.

**Analyses & Challenges Faced.** For these data, we have aligned all of the Camtasia screen video files, totaling about 50 hours, to the events in the usage log data. Using the STREAMS tools, this took approximately 30 minutes of human input.

From these linked data streams, we were able to use quantitative analysis of the usage log data to target specific events for Camtasia screen video analysis. For example, we used this method to understand sources of students' conceptual struggles by target-extracting video segments pertaining to problem steps with unusually high error rates [4].

We have also begun to work on transcribing the audio data recorded by Camtasia. However, we discovered in this process that the quality of Camtasia's audio recordings makes transcription very difficult, due to the background noise of all dyads' collaborative dialogue within a single classroom.

So, although deploying Camtasia required nearly no additional equipment and a less invasive experience for students, if recording audio in a noisy environment, it can require significantly more human effort on the analysis side.

## 3 The STREAMS Tool

The Structured TRansactional Event Analysis of Multimodal Streams (STREAMS) tool temporally aligns software-logged data files with multimodal data streams of students' learning environments. It then allows for (1) event-based extraction of relevant segments of video data, and (2) integration with DataVyu freeware for visualizing the synchronized data streams and adding new annotations.

The first component of STREAMS accomplishes temporal alignment, where different multimodal streams of data (video, audio, etc.) can be temporally synced with log data and, consequently, to each other. It uses the relative times between log data events, combined with the temporal offset between the logged data and the beginning of each media stream, to do this. If the temporal offset is not automatically recorded during data collection, then minimal human input is required to provide the time within each media stream at which the first software-logged event occurs. The output of temporal alignment is a data frame that contains the original log data plus three additional columns per synced media stream: the corresponding media stream's filename, the start time of the event within that stream, and the end time of the event within that stream. Once the data streams of interest are temporally aligned, one can either extract of segments of audio/video pertaining to specific events tagged in the log data, or visualize the synchronized streams of data for exploratory analyses and to create additional annotations.

In the event extraction component of the tool, the user can query any value of any column from the software-logged data (e.g., all problem steps tagged with skill X) or any combination of column values (e.g., all problem steps tagged with skill X on which the student made an incorrect first attempt). STREAMS will then produce a folder of extracted video segments that correspond specifically to the events specified in that query.

Finally, the tool can generate a plugin to DataVyu [3], a freeware tool that allows different data streams (including audio, video, physiology, eye tracking, motion tracking, and text annotation) to be synced in a manner that allows for easy exploratory analyses and additional annotations across streams. As it exists, DataVyu requires users to manually enter annotations. The STREAMS plugin can, however, extract data from any number of desired log data columns and automatically annotate the multimodal streams with this information within DataVyu. The result is a temporally synchronized collection of both text and multimodal data streams within an interface where additional annotations are easy to create.

A remaining goal we'd like to incorporate into the STREAMS tool is the integration of additional video annotations with the original log data. This would allow for quantitative analyses that relate data in the usage logs with multimodal data collected outside of the usage logs.

## 4 Discussion: Challenges and Future Directions in Multimodal Learning Analytics

There are benefits and drawbacks of different methods of collecting audio/video data (using individual microphones and cameras vs. computer-based screen/webcam videos). In general, if high fidelity dialogue transcription is desired, the deployment of individual microphones is important – especially when dialogue is occurring in noisy environments. However, the use of external equipment such as microphones and video cameras requires significantly higher cost and deployment effort, including the assistance and trust of students themselves to operate the equipment for recording. Camtasia is much less invasive and easier to deploy in the classroom once it's set up on the computers, but still costly beyond the free trial period. One freeware alternative may be to use Open Broadcaster Software, a method we are currently exploring.

On the data processing and analysis end, the main challenge continues to be to reduce the human hands-on time required to synchronize across the different data streams and to reduce the amount of time needed to code multimodal data while still leveraging its unique contributions. The tools we developed have alleviated some of these challenges, but there is much more to be done.

## References

1. Paulo Blikstein. 2013. Multimodal learning analytics. In Proceedings of the 3rd international conference on learning analytics and knowledge (LAK ’13), 102-106.
2. Scott Crossley, Ran Liu, and Danielle McNamara. (2017). Predicting math performance using natural language processing tools. Proceedings of the 7th international conference on learning analytics and knowledge (LAK ’17).
3. Datavyu Team. (2014). Datavyu: A Video Coding Tool. Databrary Project, New York University. http://datavyu.org.
4. Ran Liu, Jodi Davenport, and John Stamper. 2016. Beyond Log Files: Using Multi-Modal Data Streams Towards Data-Driven KC Model Improvement. In Proceedings of the 9th International Conference on Educational Data Mining (EDM ’16).
5. Jennifer K. Olsen, Daniel M. Belenky, Vincent Aleven, and Nikol Rummel. 2014. Using an intelligent tutoring system to support collaborative as well as individual learning. In Proceedings of the 12th International Conference on Intelligent Tutoring Systems (ITS ’14), 134-143.
6. Marcelo Worsley. 2012. Multimodal learning analytics: enabling the future of learning through multimodal data analysis and interfaces. In Proceedings of the 14th ACM International Conference on Multimodal interaction (ICMI ’12), 353-356.