

# TOCE vs. VLSS Submission Scoping

Austin Cory Bart

## Abstract

*A brief write-up on the tactical decisions relating to the two potential journal papers I'm working on.*

## 1. Over-arching Question

We are definitely submitting to the TOCE Learning Analytics journal; should we also submit to the VLSS Blocks journal?

## 2. Due Dates

VLSS Abstract is due March 30, final submission is due April 11.

TOCE Abstract is due May 15, final submission is due September 15.

## 3. Data Sources

The following section is devoted to explaining the different datasets.

### 3.1. BlockPy Data Interaction Logs

This is the heaviest data source. I'm estimating about 179,920 events over two-three weeks, totaling 54 MB of data. This stream is a list of events. All of the events are matched to a student (fully identified with other data from the course), assignment (including description and marked feedback), and a time.

The events can be one of two kinds:

**Code** Submission of code (blocks are represented as text, text is at the keystroke level).

**Interface event** Clicking a button, e.g., switching blocks to text, undo/redo, viewing the pseudo-code explanation, browsing through code using the explorer.

I'm using the feedback data for evaluating/guiding students to instead assess the code at every Code Submission event. The feedback consumes students' code, a trace of their variables, and the printed output, in order to return a string indicating a syntax error, run-time error, incorrect answer, or a correct answer. Although time-consuming to calculate, this gives a deep analysis of when exactly students achieve the right answer and how much progress they are making.

### 3.2. Canvas Analytics

Canvas exposes some information about its students. The API has a feature for collecting:

**Binned Page Views** - for the entire duration of the course, students number of page views for an hour is recorded. Unfortunately, this doesn't tell you what pages they were looking at, just sheer quantity.

**Participations** - the log of how many times they submitted to a Quiz or non-LTI Assignment, represented as the URL.

**Assignment Performance** - the log of a student's grade for a given assignment, including when they submitted and their score relative to the rest of the classes (min, max, median).

As far as I can tell, the Participations are overshadowed by the Assignment Performance data.

I'm investigating whether it's possible to get finer-grained data. The Canvas site itself allows me to gather data for a student about the number of times they've visited a specific page. This could be scraped, although we could also just talk to the Canvas admins about getting a dump of their user event log database, which would give exact date-times for every students access for every page. This would enable us to answer questions such as, "did students check relevant Pages on material when working on a quiz or an assignment".

### 3.3. Motivational Surveys

At the start of the semester, I administered a survey on students' motivation. This same survey will be administered at the end of the semester. If you've read my prelim, this probably sounds pretty familiar.

The first section of the survey is quantitative: 7-point likert questions in the five MUSIC categories (eMpowerment, Usefulness, Success, Interest, Caring) in relation to five course components (learning abstraction, learning algorithms, learning programming, working in a cohort, working with real-world data), for a grand total of 25 statements. There is one additional 7-point likert about their intent to follow-up on this learning experience to learn more about computing. The second section of the survey is qualitative: five open-ended questions (one for each element of the MUSIC model) about whatever course elements are most positive and negative. Finally, the last section is simple demographic data, so we can tie student responses to the rest of the data; we capture gender, major, year, and prior computing experience.

### 3.4. Canvas Quiz Grades

This semester we administered two pre/post quizzes, one on Abstraction and one on Iteration (centered around activities on the relevant subject). We have the results of these quizzes, allowing us to look at the validity of the questions, how long the quizzes took, and the growth of the student.

### 3.5. CORGIS Interaction Data

The CORGIS site is now configured to record interaction data. An interaction is recorded for loading a page, making a search/filter in the main page, and every five seconds on any page. This should help give some details on how students are browsing the CORGIS website.

Unfortunately, this system is rather ad-hoc, so it's tied to IP addresses instead of student IDs. I'm recording that IP address information elsewhere, so it might be possible to connect some student behavior with other data, but it is not as given like the other data sources.

Students are asked several times to review the CORGIS Datasets collection, and have to choose or suggest datasets for themselves to choose. This qualitative data is tied to actual student IDs.

### 3.6. Final Project

The complete codebase for every student project, along with their project video, will be available. This semester, we're using a new, more detailed rubric for evaluating students, which might be a richer source of data for evaluating projects. I'm also working on an assignment where

students have to explain automatically-selected portions of their code (e.g., a For loop is chosen, the student writes a one sentence explanation of what it is doing, "This walks through the list of earthquakes one at a time"). This can serve as a "final assessment".

### 3.7. Attendance

We have logs of students attendance throughout the course.

### 3.8. TA Evaluations

In past semesters, I asked TAs to rate their students on two dimensions on 4-point scales: general course ability and general course engagement. Although very coarse, this seemed like an interesting potential source of data. I might conceivably want to change it from a 4-point to a 7-point scale like on the motivation surveys, but I worry that that is too fine-grained for the TAs to make.

## 4. Relevant Topics from the Call for Proposals

In this section, I review the potential topics from the two call for proposals.

### 4.1. VLSS (Blocks)

The following topics are **relevant**:

- Design choices in blocks languages that help or hinder usability and learnability.
- Comparing visual and textual modalities for learning programming concepts and computational thinking, especially in the context of environments that support converting between blocks-based and text-based representations of the same program.
- Relating features of blocks programming environments to structured editors and integrated development environments.
- Disentangling the visual aspects of blocks environments from design aspects of the underlying (often domain-specific) programming languages, web-based aspects, and community-based aspects (e.g., commenting on and remixing programs shared in an online gallery).
- Issues in transitioning from blocks languages to traditional text-based languages.

The following topics **might be relevant**:

- The common perception that blocks languages are not "real" programming, and the extent to which such claims are or are not justified.

The following topics are probably **not relevant**:

- Making blocks languages more accessible to everyone, especially those with disabilities.
- Blocks programming environments that can be effectively used on devices with limited screen space, such as smartphones and tablets.

#### 4.2. TOCE (Learning Analytics)

The following topics are **currently relevant**:

- Automated feedback and grading tools and systems
- Learning Analytics in Introductory Programming Environments
- Learning Analytics to study/impact learner affect and socio-emotional factors

The following topics **might eventually be relevant** over the course of the summer development:

- Adaptive learning/personalization of content and interactions
- Adaptive tutoring systems
- Individuated scaffolds and feedback to student and instructors

I am **unsure** if the following topics could be made relevant:

- Multimodal learning analytics
- Learning Analytics and Learning Theory

The following potential topics are probably **not relevant**:

- Approaches to logging and parsing data
- Education Data Mining (EDM) software
- Implementation of Education Data Mining
- Education Data Mining methodology and Human Computer Interaction (HCI)
- Education Data Mining and data visualization
- Innovative approaches for ground truthing

## 5. Conclusion

Because the abstract is not due until May, and the final submission is open till September, I think that it makes sense to make a stab at the VLSS paper. Even if nothing comes of it, I have all of April to draft the TOCE abstract, and then all of summer to write the complete paper itself.

Crucially, some of the valuable end-of-semester data will be gathered very close to the TOCE abstract, which may affect the story that we want to tell with it.

The bulk of the VLSS paper will be analyzing design decisions in BlockPy. I will only use interaction data from the BlockPy logs where it relates to the blocks feature, rather than looking at the rich data associated with the automatic, interactive feedback. So there will be a reduced emphasis on data analysis in this paper in favor of more philosophical discussion of features.

The bulk of the TOCE paper can be comparisons of the learning analytics with an intent of being able to:

1. Identify warning signs of students early on who need interventions and attention.
2. Assess if students have performed well at the end of the course by comparing metrics.
3. Identify common failure patterns in students behavior.
4. Describe students behavior with respect to various technological course components (e.g., the BlockPy programming environment, CORGIS, Canvas).

I am not sure whether it is appropriate to make this paper be so strongly centered on the Computational Thinking class, but that's where my data is for better or worse.