Using EventFlow and CoCo to explore classroom activity patterns and learner performance

Leah P. Macfadyen
Faculty of Arts, The University of British Columbia
Buchanan C110,1866 Main Mall, Vancouver, BC, Canada
+1 604 809 5013, leah.macfadyen@ubc.ca

ABSTRACT
The ‘Classroom Observation Protocol for Undergraduate STEM’ (COPUS) tool allows observers to generate rich sets of temporal data describing in-class instructor and learner activity, but to data analysis of these data sets has been limited to aggregate reporting. Use of the temporal sequence analysis tools EventFlow and CoCo offers the possibility of exploring the temporal and sequential nature of these data sets, and uncovering relationships between classroom activity patterns and student achievement. This paper reports some sample findings from an exploratory use case of these tools.

CCS Concepts
• Visualization→ Visualization systems and tools • Visualization →Visualization application domains→ Visual analytics • Probability and statistics→ Nonparametric statistics

Keywords
Temporal analysis; COPUS; Active learning.

1. INTRODUCTION
The present SoLAR’s definition of learning analytics notes that LA may include study of “the environments in which learning occurs” [1]. In parallel, education researchers have increasingly reported that students learn more in courses that use active-engagement approaches to instruction. It remains challenging, however, to determine the degree to which instructors in higher education are employing active-engagement approaches in their teaching. This lack of data hinders institutional efforts to transform teaching and improve student learning [2].

With the goal of supporting classroom observation projects in participating institutions, Smith et al. [2] developed and published an observation protocol that would allow researchers to “reliably characterize how students and instructors were spending their time in undergraduate STEM classrooms” (p.619). The ‘Classroom Observation Protocol for Undergraduate STEM’(COPUS) is a structured protocol that offers observers a common set of 25 activity codes with which to report instructor and student activity in 2-minute intervals throughout a class session. Classroom observations using COPUS thus generate a rich corpus of temporal sequence data for each class observed, typically gathered as a matrix of co-occurring and sequential instructor and student activities (Figure 1).

	1. Students doing													2. Instructor doing												
min	L	Ind	CG	WG	OG	AnQ	SQ	WC	Prd	SP	TQ	W	O	Lec	RtW	FUp	PQ	CQ	AnQ	MG	1o1	D/V	Adm	W	O	Comments:
0 - 2																										
2-4																										
4-6																										

Figure 1. An excerpt of the COPUS coding form (from [2])

Sadly, analysis of these rich data sets so far appears to have been limited to aggregation of counts of minutes per activity, and presentation of data using pie charts, bar charts and histograms (see for example Figure 2).

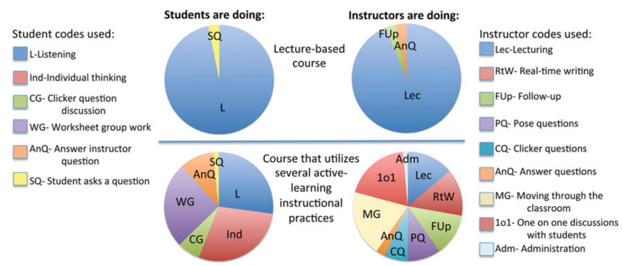

Figure 2. Pie charts of aggregate activity data collected using COPUS (from [2])

In other words, analyses published to date have failed to explore potential insights into teaching and learning in these classrooms that might be gleaned from the *temporal* nature of the data sets.

2. TEMPORAL ANALYSIS OF COPUS DATA
I am testing use of two temporal data analysis tools – EventFlow[1] and CoCo[2] – to explore data collected using the COPUS protocol in selected courses at the University of British Columbia.

2.1 EventFlow
Originally developed to allow visual analysis of temporal data from medical settings, EventFlow was developed by Dr Catherine Plaisant, Dr Ben Shneiderman and team at the University of Maryland’s Human Computer Interaction Lab[3]. EventFlow offers a timeline window which can display point and interval events for each record, while the overview window computes and visualizes an aggregated view of all the sequences in the dataset. The development team write: “Multiple graphical search capabilities enable users to find records that exhibit specific temporal patterns. Interactive features allow users to review patterns, find anomalies or construct cohorts, but also allow them to simplify the data in order to sharpen the analytic focus and answer questions about the data.” They have subsequently published extensive advice on managing volume and heterogeneity in temporal data analysis using EventFlow [3].

2.2 CoCo
Malik et al. [4] point out that visual analytic tools capitalize on the power of human visual cognition, which allows us to discern context, patterns and anomalies in large sets of data. However, visual analytic tools such as EventFlow which support exploratory and open-ended visual analysis are typically unable to also offer

[1] http://www.cs.umd.edu/projects/hcil/eventflow/

[2] http://hcil.umd.edu/coco/

[3] http://www.cs.umd.edu/hcil/

reliable descriptive statistics and automated statistical analysis. Also developed by the HCIL team, CoCo seeks to bridge this gap. It is a visual analytics tool that implements non-parametric tests to allow pairwise statistical comparison of sets of temporal sequence data. % prevalence of events and attribute significances are calculated using Chi-squared tests; time significance metrics use a Wilcoxon sum-rank test across the distribution of values [4]. Results (comparative statistics for measures such as prevalence and frequency) are displayed visually.

2.3 The data

Our sample data set comprises observational records collected using the COPUS protocol in 31 separate classroom observations of 15 different biology courses. Courses range from 1st to 4rd year, were taught by 26 different instructors, and varied in class size. Importantly, all made use of diagnostic tests (pre-tests and post-tests) which allowed computation of average 'learning gain' per class cohort, and relative 'effect size' of this learning gain.

3. SAMPLE FINDINGS

This study is exploratory, and is a use case work in progress, in consultation with the U. Maryland HCIL development team. In first round analysis I used EventFlow to visualize patterns of instructor and student incremental activity data as 'point events', and made use of activity codes available in the COPUS data set. I have begun to undertake comparative visual analysis of courses where learners achieved on average top quartile or bottom quartile learning gain. Figure 3, for example, shows patterns of 'instructor lecturing' (purple), 'students listening' (yellow) and 'students doing independent work' (brown) in two different courses: course 34 (three observations), and course 12 (four observations). Notably, students in course 12 achieved 'top quartile' learning gain on average, while the course 34 cohort place in the bottom quartile for learning gain.

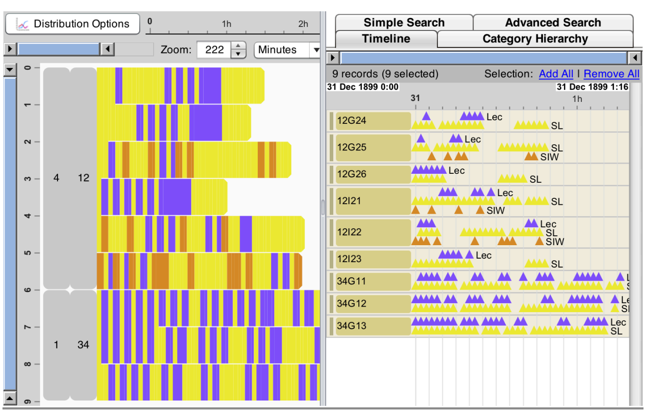

Figure 3. Sample EventFlow visual analysis of COPUS classroom observation data.

CoCo, allows a statistical comparison of activity sequences in classes placing in the top ("4") and bottom ("1") quartiles of learning gain. Figure 4 shows sample CoCo output. At first glance, this visual representation of comparative sequence analysis indicates, for example, that the frequency of use of clicker questions (CQ, blue) and moments of independent student work (brown, SIW) are significantly higher in top quartile-achieving courses. A more detailed discussion of findings will be presented later. In ongoing work, I am continuing to refine event categories, and to create intervals from point data, to permit more meaningful characterization of course 'types'.

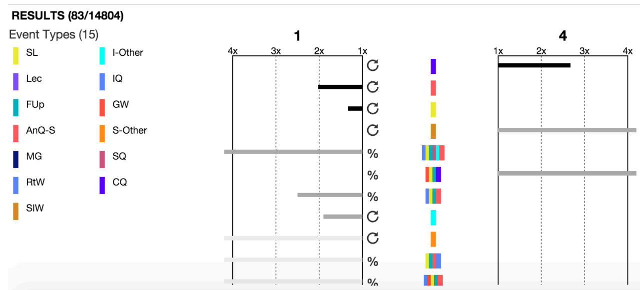

Figure 4. Sample CoCo statistical/visual analysis of COPUS-generated sequence data

4. ACKNOWLEDGMENTS

I am grateful to Dr Ben Schneiderman and Dr Catherine Plaisant of the HCIL team at U. Maryland for their willingness to allow exploratory use of their analytic tools, and to team members Sana Malik and Fan Du for ongoing consultation and support.

5. REFERENCES

[1] SoLAR. (2011). *Open Learning Analytics: An Integrated & Modularized Platform. White Paper*. Society for Learning Analytics Research. Retrieved from http://solaresearch.org/OpenLearningAnalytics.pdf

[2] Smith, M. K., Jones, F. H. M, Gilbert, S. L. and Wieman, C. E. The Classroom Observation Protocol for Undergraduate STEM (COPUS): A New Instrument to Characterize University STEM Classroom Practices. *CBE-Life Sciences Education* 12 (Winter 2013), 618–627.

[3] Shneiderman, B. and Plaisant, C. Sharpening Analytic Focus to Cope with Big Data Volume and Variety. *Computer Graphics and Applications, IEEE* 35,3 (May-June 2015), 10-14. DOI=10.1109/MCG.2015.64

[4] Malik, S., Du, F., Monroe, M., Onukwugha, E., Plaisant, C., and Shneiderman, B. Cohort comparison of event sequences with balanced integration of visual analytics and statistics. In Proceedings of the 20th International Conference on Intelligent User Interfaces (2015). IUI '15. ACM, New York, NY, USA,38-49. DOI=10.1145/2678025.2701407