Do You Speak Math? Visualizing Patterns of Student Technical Language in a Mathematics MOOC

Paul Franz, Brian Perone, Stanford University, 485 Lasuen Mall, Stanford, CA 94305
Email: pefranz@stanford.edu, bperone@stanford.edu

Abstract: The recent advent of massive open online courses (MOOCs) has created an incredibly vast, rich, and largely unexplored body of data. We have analyzed one MOOC by visualizing the language used by students in their online forums, connecting vocabulary to traditional performance assessments throughout the course, prototyping a tool for future use across courses and platforms.

Background

Learning is a dynamic process of interactions between learners, resources, environments, and instructors (NRC, 2000). In a massive open online course (MOOC), these interactions take place within a distributed learning environment, bounded by the decisions made by faculty around the design of instruction, content, and assessment, as well as platform technological features and the incredible range of student backgrounds and intentions (Grover, Franz, Schneider, Pea, 2013). A key challenge of this new space is assessing learning – researchers have yet to establish definitive metrics about which data will provide the most insight into how people learn in MOOCs, and data useful for researchers may not be the same as informative data for instructors or for individual students. In addition to commonly used assessments of performance, such as exams and attitudinal surveys, learning in a technical field, such as mathematics, can be viewed as the acquisition and use of professional or "academic" language (Pea, 1990; Van Oers, 2000). Successful mathematics education of any kind – online, face-to-face or blended – should result in capabilities to participate in mathematical discourse. Most importantly, students must be able to engage in mathematical communication and argumentation with their peer community. "The meaning of representations such as words and diagrams in a community becomes evident through their use and the reshapings of their meanings through commentary by other participants of learning conversations" (Pea, 1990). The course forums provide a clear record of interactions within exactly such a peer community.

Visualization can serve as a powerful approach for exploring and making sense of the complex and multidimensional set of performance data (Bienkowski, 2012), including the connections between multiple dimensions of performance. With this in mind, we have developed a prototype for the visual exploration of multidimensional features of student learning in MOOCs.

Forum and Survey Data: Research in Progress

The specific question that the visualization is designed to help answer is: how does the adoption of technical, academic language in the *Introduction to Mathematical Thinking* MOOC on Coursera correspond to instructor language use and student performance on course assessments? More specifically, when do students begin to use formal academic language introduced by the instructor, and does that timing correspond to their performance on other, more traditional, performance metrics?

The *Introduction to Mathematical Thinking* MOOC was a 7-week course taught in Fall 2012. As part of the course, students had access to a discussion board where they could socialize, seek help, and explore mathematical concepts. The course began with 44,432 active students. Of these, 5,066 posted in the forum at least once for a total of 3,235 threads consisting of 33,828 individual posts. Of the forum participants, 1,214 completed the course pre- and post-survey, and this group forms our population of study in this visualization.

The visualization has three parts. First, individual student scores on problem sets are encoded in a parallel coordinates graph, with score as a percentage indicated by vertical position. This graph also contains pre- and post- class survey responses to the question, "In general, how relevant to you are the things that are taught in mathematics classes?" Lines for students who never used the currently selected word are red, so they can be distinguished from lines for students who did use the word, in blue. Data for specific students or groups of students can be selected, and their trajectories through the course assessments will be highlighted. Below that, a timeline shows the median time of first forum use of the word for the students selected in the top graph, as compared to the time of the instructor's introduction of the word. Finally, a histogram shows all use of the word over the entire course. Live visualizations for several words of interest are available at www.stanford.edu/~bperone/wordUseVisualization.

The primary goal of this visualization tool is to aid in exploratory data analysis (Tukey, 1977) by highlighting trends that can be pursued in more depth using other analytical tools. It can also be adapted to answer similar questions around language adoption and performance in other MOOCs. Finally, it provides a prototype for real-time exploration of the interactions between multiple types of performance data.
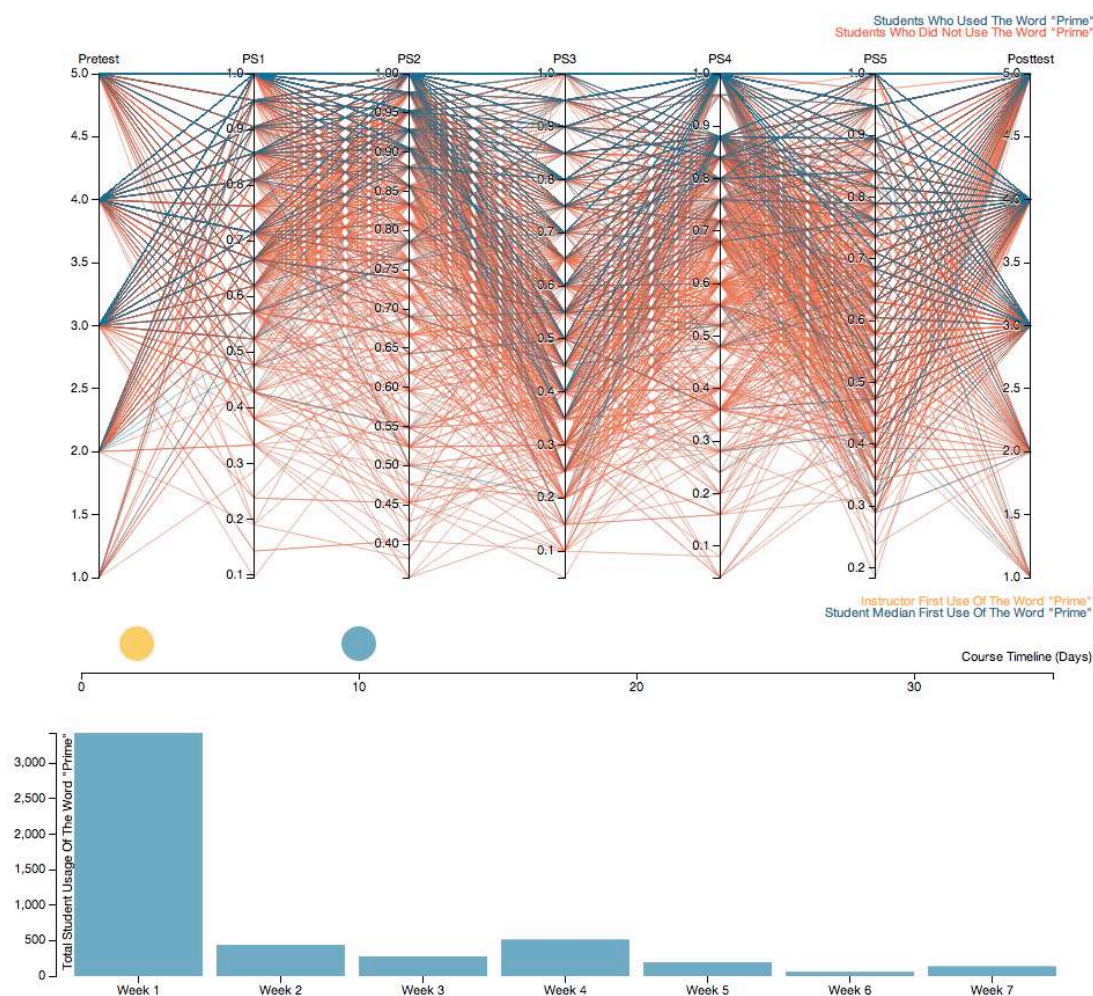
Figure 1. A screenshot of the visualization

Future Work

An obvious limitation of our approach is the lack of context for word use. Future work could use more sophisticated natural language processing to examine not only when words are used, but how. For example, we could use markers for exploratory language (Ferguson, Buckingham Shum, 2013), or determine whether key words are used declaratively or interrogatively. Separating out individual forum threads and looking at the conversational interactions around specific topics could also prove illuminating.

References

Bienkowski, M., Feng, M., & Means, B. (2012). *Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief*. U.S. Department of Education Office of Educational Technology.

Ferguson, R., Wei, Z., He, Y., & Buckingham Shum, S. (2013). An evaluation of learning analytics to identify exploratory dialogue in online discussions.

Grover, S., Franz, P., Schneider, E., Pea, R. (2013). The MOOC as distributed intelligence: dimensions of a framework for design & evaluation of MOOCs.

National Research Council. (2000). *How People Learn: Systematizing and Generalizing the Tool Brain, Mind, Experience, and School: Expanded Edition*. National Academies Press.

Pea, R. D. (1990). Augmenting the discourse of learning with computer-based learning environments. In *Proceedings of the NATO Advanced Research Workshop on Computer-Based Learning Environments and Problem Solving, Held in Leuven, Belgium, September 26-29, 1990* (pp. 313–343).

Tukey, J. W. (1977). Exploratory data analysis. *Reading, MA*, *231*.

Van Oers, B. (2000). The appropriation of mathematical symbols: A psychosemiotic approach to mathematics learning. *Symbolizing and communicating in mathematics classrooms: Perspectives on discourse, tools, and instructional design*, 133–176.