ABSTRACT
Researchers in the developing field of Learning Analytics need new visualization tools to support the analysis of data sets from online learning platforms such as Massive Open Online Courses (MOOCs). The multidimensionality of data collected from MOOCs creates opportunities to explore and model the complex psychological and social process of learning. One area in need of visualization work is language use by learners in online environments, and how that language use corresponds to student performance and attitudes towards the content domain. In this paper, we describe the initial design and prototyping of a tool to help researchers address these questions, using data from the Introduction to Mathematical Thinking MOOC on Coursera.

Author Keywords
Learning analytics, academic language, multiple data representations, massive open online courses.

ACM Classification Keywords
H5.2. Information interfaces and presentation (e.g., HCI): User Interfaces.

INTRODUCTION
Education is becoming increasingly digitized. Virtual environments from MOOCs to the Khan Academy to educational software in classrooms are recording each move that a learner makes. Learning analytics is developing as a field to harness this influx of “big data” - the goal of the field is to develop strategies for the measurement, collection, and analysis of data, and to use these analyses to improve the design of learning environments [2]. What are the meaningful signals of learning among this data? How can we represent and understand the complexity of learning by visualizing multiple dimensions of learner activity and characteristics?

Learning is a dynamic process of interactions between learners, resources, environments, and instructors [5]. In a MOOC, these interactions take place within an environment bounded by the decisions made by faculty around the design of instruction, content, and assessment, as well as the technological features of the platform. The pathway that each learner takes through this space is a personal one, informed by her own background, intentions, and prior knowledge, as well as by her experience in the course as time goes on. Examining these pathways in the aggregate provides insights into overall trends in learning.

In principle, the fine-grained interaction data logged from MOOCs allows us to model this complex psychological and social phenomenon. There remains much work to be done, however, in selecting, representing, interpreting, and analyzing data from MOOCs. Researchers have yet to establish definitive metrics about which data will provide the most insight into how people learn, and the data that is useful for researchers may not be the same as data that is informative for instructors or for individual students. For each of these groups, visualization will serve as a powerful approach for exploring and making sense of the complex and multidimensional set of available data [1]. With this in mind, we develop a prototype for the visual exploration of multidimensional features of student learning in MOOCs.

TARGET PROBLEM
The specific question that our 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? Research in the learning sciences shows that the development of domain-specific, academic language is a powerful indicator of learning [6], particularly in mathematics [8]. Moreover, in the course from which data for this prototype was selected - the Introduction to Mathematical Thinking MOOC - one of the explicitly stated desired learning outcomes for students is the ability to think and speak precisely about mathematical topics.

The Mathematical Thinking MOOC was a 5-week course. Each week a new set of lecture videos were posted, and a new problem set was assigned. Students had access to a discussion board where they could socialize, ask for help, and explore mathematical concepts. The discussion board was monitored by the instructor and a teaching assistant at Stanford, as well as a set of volunteer “Community TAs.” 44,432 were active students at the beginning of the course, and 5,066 of these students were active on the forum.
We narrow in on the learning process of academic language adoption by comparing the timing of the instructor’s usage of a term to its adoption by students, and characterize students by their performance on assessments and their attitudes toward math. With assessments, learners are demonstrating their command of the concepts at hand - we can look at problem set scores and see whether someone is struggling with the topic or has mastered it. With social interactions on a discussion board, learners are demonstrating their understanding of technical or academic language.

The visualization tool we have developed aids in exploratory data analysis by highlighting trends that can be pursued in more thorough statistical or natural language analyses. It can also be adopted to answer similar questions around language use and performance in any number of MOOCs. Finally, it provides a prototype for real-time exploration of the interactions between multiple types of performance data.

RELATED WORK
Within the broader domain of Learning Analytics research, and in particular into MOOCs, there are a large set of possible design tasks that are in need of good data visualizations. MOOCs have emerged as an educational phenomenon with the last year, and as a result, there has been no published work on evaluating learning in these environments, let alone using visualization as a strategy for exploratory data analysis. Visualizing learning data is also a relatively new endeavor. In the limited literature available, the visualization strategies chosen follow from the underlying data and the types of information arising from this data that would be most beneficial for a researcher, instructor, or student. For example, intelligent tutoring systems lead students through a personalized pathway of exercises on a particular topic, based on an underlying map of the knowledge space. Visualizations of these pathways are represented as node-link diagrams or hierarchical tree diagrams [3, 7]. The Khan Academy’s website follows a similar model for their “Knowledge Map” through the available exercises.

The closest example to our current work is CourseVis, a tool designed to represent student participation in discussion boards and assessments in distance learning courses – distinct from MOOCs in that these courses have limited enrollments and a low instructor to student ratio [4]. CourseVis represents multiple dimensions of performance for an individual student by date, and also provides an aggregate visualization of discussion board activity by date. These representations are useful for an instructor hoping to understand the performance trends of a particular student in order to provide personalized support, or to get a sense of activity in the course as a whole - for example, are there particular dates on the message board where there was a spike in activity? What topics were being covered in the course at that time?

Compared to CourseVis, a unique contribution of our work is the relation of aggregate trends in student achievement and attitudes to other signals of learning - namely, academic language use. Our tool also builds the foundation for making a contribution to the literature on academic language use. In contrast to prior work relying on transcripts from classroom discussions [8], we benefit from the corpus of data available on the discussion board.

DESIGN REQUIREMENTS
In order to support learning analytics researchers looking at language use and performance, we drafted the following design specifications. Not all of these specifications have yet been implemented, but they nevertheless reflect our vision for what we are trying to build. Not yet implemented aspects of the visualization appear below in italics.

- We should display student performance using available data, selecting those data most relevant to learning and engagement.
- We should include data on language use, both aggregated and sampled for the sake of comparison.
- We should pre-select important technical words, as well as allowing users to query words of interest.
- We should characterize word usage as accurate, inquisitive, misconceived, or some other qualifier. We should allow the user to filter based on these qualifiers when they search.
- We should support comparison of language use and performance in a single display. Our hypothesis is that there is a connection, but that it is a complex one.
- We should allow exploration through interactive features that enable the selection and sampling of students across any variable in the data.

VISUAL ENCODINGS AND INTERACTIONS
Figures 1 and 2 (following page) show our final visualization of the word “Prime,” first without any interaction and second after the user has selected various parameters. Figure 3 shows the same visualization of the word “Proof.”

The visualization encompasses three dimensions of learning in the MOOC. Two of the dimensions reflect the learning pathways of individual students and the third provides context for the MOOC as a whole. From the top to the bottom of the visualization, these are:

- **Learning dimension:** Student-level performance and attitude data
- **Learning dimension:** A course timeline that indicates the instructor’s first usage of the term of interest, as well as the median time of first use of the term.

1 Currently we support selection on performance data, but not yet on language use.
• **Contextual dimension:** Total usage of the term of interest across all participants in the discussion board, mapped to the course timeline.

**Exploring Academic Language Development in Massive Open Online Courses**

Connecting multiple dimensions of performance to adoption of mathematical terminology

**Figure 1** - The entire data display for the word "Prime."

**Figure 2** - Selecting a portion of the data recalculates the median first usage moves based on the current selection.

**Exploring Academic Language Development in Massive Open Online Courses**

Connecting multiple dimensions of performance to adoption of mathematical terminology

**Figure 3** - The entire data display for the word "Proof."

**First Learning Dimension: Assessment Scores**

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?” This question was answered on a 1-5 Likert scale, with 5 as “Extremely Relevant” and 1 as “Not at all Relevant.”

Each axis is scaled so that the lowest score on the graph is the lowest score received on the assignment. This normalization enables a rank-ordered comparison across axes. Lines between points on the axes allow the user to connect an individual student’s scores across all assessments, as well as indicating change in performance from one activity to another. The encoding by vertical position captures an instinctual interpretation of “high achievers” and “low achievers”, and allows the viewer to easily compare students’ achievement level at different points in the course. The default position of the axes reflects a time-series - students took the pre-survey, then did each problem set, and ended with the post-survey - but the axes can be rearranged for closer comparisons between different scores. Putting the pre- and post-survey next to each other allows easy detection of the students for whom reported math relevance changed over time.

To dig deeper into the categories of students, users can filter on achievement level or mathematical relevance. Users select a sample of students by clicking and dragging a selection box around any data set of scores on any Problem Set or Survey, or on multiple Problem Sets or Surveys. For example, a user might select all students who answered...
“Extremely Relevant” on both the pre- and post-class surveys, or they might select students who received 75% or higher on the first Problem Set and lower than 50% on the second. These selections will highlight selected users on the parallel coordinates graph, as well as changing the data display in the second graph.

Lines for students who never used the currently selected word are red, so they can be distinguished from students who did use the word, in blue. Lack of word does not necessarily mean that the student did not understand the word; rather, she may not have been an active forum participant, despite completing all the assignments and surveys.

Second Learning Dimension: Word Usage Timeline
The timeline uses a symbol, encoded by color, to indicate the first usage of the currently selected word by the instructor and the median first usage by the current sample (from among those who did use the word). The user can thus probe the relationship between different kinds of scores on different Problem Sets and survey responses and the use of technical mathematical language. By displaying the timeline in days, the user can get a fine-grained sense of when a word was ‘adopted’ by a target sample of the student population, relative to when that word was introduced.

Contextual Dimension: Aggregate Word Usage
The bottom graph displays total number of uses of the word by all students in the course in the forums, sorted by week. This graph allows the user to compare the median adoption by the current sample to overall trends in word use. We use a histogram to show word use over time because it a simple representation of frequency that allows for straightforward comparison between weeks. Some of the trends are driven by overall levels of activity on the discussion board, with more students active at the beginning of the course than the end. There are a greater number of students active in the course at the beginning than the end, so some of the trends in word use are driven by this. Once dynamic word queries are implemented, this will also allow users to track how different words are used over time. Comparing to instructor usage is particularly informative here, as some words are used from the outset of the course, well before the instructor’s first use, while others peak sharply in the week of introduction.

Linking the Graphs
Each of the three graphs produced by our visualization contains relevant data, but we see the primary contribution and innovation of our system as the simultaneous and interconnected display of all three. The parallel coordinates and timeline allow us to first de-aggregate a large dataset, and then to re-aggregate it across selected students. The bottom graph shows an even higher level aggregate view of the data, comparing the sample of students in the first two graphs with students from the entire course. This simultaneous display of three linked graphs, each at a different level of granularity, is particularly valuable for a learning analytics researcher who wishes to dive more deeply into the relationship between performance and language use by students in the course without losing track of the bigger picture of the data.

Furthermore, the bottom two graphs are aligned so that they are chronologically consistent moving from left to right. That is, “week one” on the bottom graph lines up with days 1-7 on the middle graph. The top graph is ordinarily aligned, but because the surveys were not administered on a time schedule, and because students could take the Problem Sets over a range of days, it is impossible to fully compare with the other graphs. Nevertheless, the parallel coordinates do roughly line up with the days of the course indicated in the middle graph. This further supports the ability to focus on different parts of the visualization by maintaining a consistent x-axis across all three graphs.

DATA TRANSFORMATIONS
In order to represent our data effectively in our design, we engaged in four data transformations that greatly simplify the underlying data. These simplifications were appropriate for developing a prototype of the tool, as a primary goal was developing solid functionality that we can build on in the future. Below we explain each data transformation and briefly discuss how each affects our design and our interpretation of the data.

Student Scores on Problem Sets
We dropped students who had missed any of the Problem Sets or surveys from the sample. This served as a proxy for engagement in the course - as we were most interested in the more engaged students - as well as solving the major design problem of how to represent missing data. As a result, however, users of our visualization will be best served to think of it as an exploratory tool, and not a final analysis.

Student Survey Responses
In addition to dropping non-responders, we also selected a particular survey question which we believe serves as a basic proxy for a background characteristic of the learner. We can hypothesize that students for whom math is more relevant may be more actively engaged or higher performers in the course. We can also look at whether self-reported math relevance increases at the end of the course, and whether there are any trends among these learners in word usage or problem set performance. However, our survey contains many other questions of potential interest. Future work on our project may include the ability for the user to select which survey question (or set of survey questions) to display in the visualization.

As a result of keeping only data for students who completed both the pre- and post-survey, as well as all five problem sets, our visualization shows data for 1,217 students.
Student Forum Posts
We decided to query the forum data for first student usage of a word, by day of the course. We are aware of the limitations of word usage as deeply meaningful metric and plan to engage in more sophisticated language processing in the future, in order to classify word count along various dimensions. For now, we queried aggregate usage data on a single word over the entire forums by week. The difference here stems from decisions we made about visual encodings: daily usage of words on the aggregate creates more noise in the data than is desirable for our objectives. While usage patterns by day are interesting - for example, there tends to be more forum use on the weekends - that is not what we are trying to demonstrate in this particular visualization. Rather, the purpose of aggregated word usage data is to give a quick sense of overall trends in usage for comparison with individual student and instructor usage.

Instructor Transcripts
Finally, we queried instructor transcripts to find the first use of a word in the video lectures. We represent the day on which the video in question was released on the visualization, which makes the individual or median student usage particularly interesting: are students using words before, right after, or long after the instructor does so?

A PROMISING USE CASE
Though this project is still very much a prototype, it can already inform further statistical analysis of the data. One of the virtues of data visualization is that it can help uncover counterintuitive results. In the case of this prototype, tracking the median first use of the word “prime” by selecting a single survey response score at a time reveals a surprising trend. We would expect students who claim mathematics is more relevant to their lives to adopt mathematical language sooner than students who find mathematics less relevant. However, students who said that math was more relevant to their day-to-day lives actually use the word “prime” later in the course than students who did not. Running a regression on this data does indeed show a small, positive, but statistically significant correlation (see figures 4 and 5). Though the relationship is not as strong, this same relationship appears in a second version of the visualization using the word “proof.”

There are several possible accounts for this surprising result, the most intuitive being that students with lower math relevance are less familiar with mathematical terminology, and thus are more likely to ask questions about technical language, while those with higher relevance may already know the material and so skip these earlier, less sophisticated uses. Using the visualization, we can see whether this relationship holds for all levels of performance in the course, as well as testing to see whether the same relationship appears to hold for other words. Developing a more fine grained sense of how words are being used is an important next step in answering the questions this result raises.

Interaction with the visualization suggests a similar relationship between performance on problem sets and word adoption (see figures 6 and 7, this time using the word “Proof”). If this holds across a variety of words, it is an especially provocative result. In combination with the relationship between word adoption and sense of relevance of mathematics and prior research on language adoption, this would suggest that the students who did “best” in this course may not have been as actively engaged in the learning process as those who did worse, but were active in trying out language on the forums and asking questions or collaborating with other learners. This has an intuitive appeal, but needs to be investigated more deeply before any claims can be made.
For the purposes of the visualization the point is not to answer these kinds of analytical questions but to rather raise them. It seems unlikely that we would have tested for either of these relationships without our visualization, or, even if we had, that it would have been as striking and noticeable as it is.

**FUTURE WORK**

There are several aspects of our visualization that we already have plans to improve.

**Systematizing and Generalizing the Tool**

We will allow users to dynamically query terms of interest by integrating our Python data analysis code more fully with D3 in a Django web environment, something we did not have sufficient prior knowledge nor learning time to do for this prototype. We also plan to make the system adaptable to a range of performance data such as different types of assessments and other indicators of engagement. This will allow us to import data from a variety of Coursera classes, or classes from other online learning platforms.

**Refining Underlying Data**

We use a relatively basic indicator of language use – timing and frequency of particular terms. For a deeper understanding of learning we will need to examine language in context – for example, whether terms are being used formally or informally, accurately or inaccurately, inquisitively or declaratively. We plan to develop significantly stronger competency in Natural Language Processing in order to do this work, and will revise our visualization system to reflect the most salient indicators of learning and engagement.

**User Testing**

As learning analytics researchers we are our own pilot group, but we plan to share the tool with other researchers to understand if it is useful for them to answer the research questions they are interested in. We will ask about their outcomes of interest and consider the meaningful ways that these outcomes could interact. As a prototype, this design is still subject to significant revision based upon the feedback we receive from our colleagues.

**ACKNOWLEDGMENTS**

We thank Keith Devlin for opening his virtual classroom to us, Jeff Heer and the TAs in CS448b for their help with code, and finally Relly Brandman, Pang Wei Koh, Chuong Do, and others at Coursera who supported the course.

**REFERENCES**


