Understanding Student Procrastination via Mixture Models

Jihyun Park
Department of Computer Science
University of California, Irvine
Irvine, California 92697
jihyunp@ics.uci.edu

Renzhe Yu
School of Education
University of California, Irvine
Irvine, California 92697
renzhey@uci.edu

Fernando Rodriguez
School of Education
University of California, Irvine
Irvine, California 92697
fernanr1@uci.edu

Rachel Baker
School of Education
University of California, Irvine
Irvine, California 92697
rachelbb@uci.edu

Padhraic Smyth
Department of Computer Science
University of California, Irvine
Irvine, California 92697
smyth@ics.uci.edu

Mark Warschauer
School of Education
University of California, Irvine
Irvine, California 92697
markw@uci.edu

ABSTRACT
Time management is crucial to success in online courses in which students can schedule their learning on a flexible basis. Procrastination is largely viewed as a failure of time management and has been linked to poorer outcomes for students. Past research has quantified the extent of students' procrastination by defining single measures directly from raw logs of student activity. In this work, we use a probabilistic mixture model to allow different types of behavioral patterns to naturally emerge from clickstream data and analyze the resulting patterns in the context of procrastination. Moreover, we extend our analysis to include measures of student regularity–how consistent the procrastinating behaviors are–and construct a composite Time Management Score (TM). Our results show that mixture modeling is able to unveil latent types of behavior, each of which is associated with a level of procrastination and its regularity. Overall, students identified as non-procrastinators tend to perform significantly better. Within non-procrastinators, higher levels of regularity signify better performance, while this may be the opposite for procrastinators.

Keywords
Procrastination, Regularity, Time Management, Student Modeling, Clickstream Data, Online Courses, Probabilistic Mixture Model, Poisson distribution

1. INTRODUCTION
As colleges and universities continue to increase the number of online course offerings, these classes are becoming a normal part of students' learning experiences. While online courses have made learning more accessible to students, prior work suggests that students enrolled in online courses have worse learning outcomes when compared to students enrolled in face-to-face courses [2]. One important reason for this is that the online learning environment requires a higher degree of self-regulation than the face-to-face environment [5]. Students must effectively plan and regulate their learning time, and monitor their own progress in order to meet important deadlines [31], but students may lack some of these important skills. Moreover, online courses have a high degree of anonymity. Students are not physically present in a classroom, and their activity on Learning Management Systems (LMS) is not made public to the rest of the class. This absence of face-to-face accountability may cause students to disengage with the course much more than they would in traditional classrooms. The lack of structure and anonymity may lead students to procrastinate, putting off work until close to important deadlines. Therefore, understanding students' learning behaviors relating to time management, especially procrastination, could be one important mechanism for improving online learning.

Clickstream data sets have provided rich resources for analyzing students' time management behaviors. Procrastination has been measured using the specific time points at which students take certain actions within an online course, such as accessing content pages, watching lectures, and submitting quizzes. A common way to measure procrastination is to calculate the amount of time a student is engaged with the LMS prior to an important course deadline. Studies that use these types of measures as indicators of procrastination find that the indicators are negatively correlated with course outcomes [14, 16, 30]. In the context of studying planning behaviors, researchers have also developed measures of student regularity are in the timing and spacing of their course activities, and found that higher measures of regularity correlate with better performance [28, 3].

Motivated by these previous studies, we utilize clickstream data to further understand procrastination using two online classes offered at a large public university. These two classes were designed so that the students are expected to space out their studies on a daily basis, and to set weekly deadlines. In this paper, we investigate the use of probabilistic mixture modeling to analyze time-stamped logs of student activity in the context of these two online classes. The mixture model identifies different behavioral patterns in the data, where the patterns can be clearly identified as reflecting procrastinating and non-procrastinating behavior among the students. Moreover, we notice that while procrastinating students may procrastinate frequently, some may also exhibit a mix of planning and procrastinating behaviors throughout the course. To capture these nuances, we construct a composite score, which incorporates both the overall degree of procrastination and the regularity of procrastinating behaviors. This score captures behavioral differences of procrastinators, a notion which has been absent in prior research. The methodology we develop enables finer-grained analysis of procrastination and its relationship with learning outcomes, which can inform more effective instructional reforms in online learning.

The primary contributions of this work are four-fold.

- First, we develop a general data-driven method for identifying procrastination. This method analyzes counts of student activity and can work with any online course with periodic deadlines and that has corresponding time-stamped clickstream data.
- Second, we validate this method using two online university classes, and identify two distinct behavioral patterns which can be used to measure an individual student's degree of procrastination.
- Third, building off of prior measures of procrastination, we investigate the regularity of procrastinating behaviors and incorporate this information into a composite score, providing a more detailed perspective on procrastination.
- Fourth, for the two classes we analyze, we find that all of our measures of procrastination are highly correlated with course outcomes, lending support to prior theories of self-regulated learning and procrastination while also providing new insights.

2. RELATED WORK

2.1 Self-Regulation, Procrastination, and Academic Success

Self-regulated learning refers to the process of directing one's own learning experience [31] and these processes encompass several attitudes and behaviors. For instance, models of self-regulated learning generally distinguish between motivational beliefs about learning, goal setting and planning behaviors, specific learning strategies, and metacognitive monitoring processes [22]. While each of these facets play an important role in the learning process, research on online learning finds that students' planning and time management behaviors are important indicators of course success [10, 30]. Procrastination behaviors, which refer to delaying coursework until major deadlines, reflect poor planning and time management.

Several studies have focused on procrastinating as a major barrier that hinders students from succeeding in online courses [10, 29]. Using online course analytic data, one recent study found that students who did not begin working on assignments until hours before a deadline received lower course grades when compared to students who began their work earlier [9]. Other studies have found similar results, where students who delay working on assignments are more likely to perform poorly [29, 30]. These results confirm the undesirable nature of procrastination as well as the importance of regular learning behaviors.

Another extensive body of work has shown that students from underrepresented backgrounds, such as those who come from low-income households, or who are first to attend college, are a greater risk for leaving STEM majors [7]. This problem may be additionally exacerbated in online coursework. There are many important factors that explain issues surrounding underrepresented student success, such as lack of mentoring, financial concerns, and feelings of exclusion [25]. With regard to self-regulatory behaviors such as procrastination, prior work has also shown an increased tendency for underrepresented groups to engage in more procrastination than the counterparts [24]. However, this study was not conducted in an educational context and the procrastination was measured subjectively using surveys. With this in mind, a side aim of our work is to explore the relationship between individual differences in procrastination (time management behavior, in general) and students' external background characteristics, specifically for the students taking online courses.

2.2 Measuring Procrastination and Regularity

Measures of procrastination are relatively straightforward and similar across various learning environments. In the most common measures, researchers capture the time that students finish a certain task and calculate the difference between this time and either the release time [3] or the deadline [16, 14] of the task. This type of measure has the merit of being very interpretable, but a limitation is that it only captures the average degree of procrastination without depicting nuanced patterns in these behaviors.

Regularity, on the other hand, is a higher-order concept that allows for different definitions. Accordingly, there has been a slightly larger pool of measures in the literature. Some studies define regular behaviors as repeating certain temporal patterns in a cyclic manner, and apply methods from signal processing to model hidden frequencies within students' behavioral streams [27, 3]. Another popular way of operationalizing regularity is to relate regularity to changes of learner behaviors, and quantify the changes via measures of variation [1, 28, 23] or explicit statistical modeling [21]. These different definitions are not exclusive and share many similar properties.

Most of the existing studies regarding time management in online learning examine either procrastination or regularity, and those that investigate both treat them as independent features of student behaviors. Our work extend these studies by understanding how regularity and procrastination are interrelated.

2.3 Cluster Analysis and Mixture Modeling

Clustering in general is a widely used technique in data analysis for automated data-driven discovery of groups or clusters in data. In the context of analyzing education data, clustering algorithms have found broad application as a technique for clustering of students into groups based on their behavioral patterns. For example, Toth et al. [26] cluster students based on their problem-solving interaction patterns using the X-means algorithm (a variation of the well-known K-means clustering algorithm) for a better understanding of complex problem solving behaviors and identifying levels of problem solving proficiency. Ng, Liu, and Wang [20] use survey scores of motivated strategies for learning questionnaires to cluster students into multiple groups. The resulting groups obtained by hierarchical clustering with Ward's method exhibit distinct learning profiles of motivational beliefs and self-regulatory strategies.

The clustering approach we follow in this paper is probabilistic model-based clustering [12, 19]. In this framework, each cluster corresponds to a probability distribution (also known as a "component") in a mixture model and the entities being clustered are assumed to have been generated by one of the component distributions. This probabilistic framework for clustering has a number of advantages over non-probabilistic techniques such as K-means clustering or hierarchical clustering. For example, as we describe later in the paper, the framework allows us to model count data in a natural manner using Poisson distributions as components in the mixture model, where each component (or cluster) represents a different Poisson distribution over count outcomes. The Poisson mixture model has been applied to a number of different fields including marketing [4], finance [15], biology and bioinformatics [6, 11], document analysis [17], and so on. However, to our knowledge, there has not been any prior work on the development of Poisson mixtures in an education context, particularly for the problem of clustering students based on their observed activity in online classes.

3. METHODS

3.1 Student Activity Counts

For courses where time-stamped student-generated events are tracked via logs of clickstream data, we can count these events on a daily basis. Thus, we can get a set of *daily activity counts* for each student throughout a course, where the activity can correspond to specific types of tasks of interest (such as video-watching, quiz submission, and so on) Figure 1 shows *daily activity count* data for 2 students from one of our course data sets. The data for each student is displayed as a matrix, where the grayscale indicates the number of tasks performed by each student on each day over the 5 week duration of the course. This type of display is useful in terms of capturing the temporal aspect of when a student is engaged in a particular activity such as watching a lecture video or submitting a quiz. It also indicates that one of the students (on the right) may be procrastinating each week—we discuss these types of patterns in more detail later in the paper.

We can also compute the sum over weeks to get the *aggregated daily counts* assuming that there is a structure in the course that repeats every week (which is the case for the two courses we study in this paper). Examples of the *aggregated daily counts* are shown in Figure 2 as bar plots, computed by aggregating across the weekly rows of data for each student in Figure 1.

3.2 Mixture Model with Gamma Priors

In this section we discuss our use of a Poisson mixture model to cluster students based on their activity counts, focusing on the *aggregated daily counts* as in Figure 1. In terms of notation we let $\boldsymbol{y_i}$ be the vector of *aggregated daily counts* for student i, where $i = 1, \ldots, N$. The dimensionality D of each vector is the number of days ($D = 6$ in this case since Saturday and Sunday are collapsed into one). Thus,

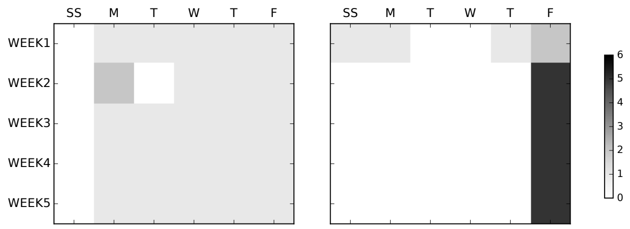

Figure 1: Examples of student *daily activity counts* (specifically, the number of video watching tasks per day) displayed as a matrix of week × day counts. SS indicates Saturday and Sunday.

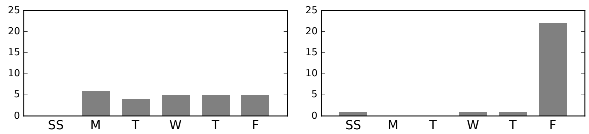

Figure 2: ***Aggregated daily task counts*** **across weeks ($\boldsymbol{y_i}$) for the two students shown in Figure 1.**

our data consists of N students each with a D-dimensional vector of *aggregated daily counts*.

To model this data we use a probabilistic mixture model with Poisson components. Let K be the number of components (or clusters) with an index $k = 1, 2, \ldots, K$. The unobserved latent variable z_i takes values from the set $\{1, \ldots, K\}$ and corresponds to the latent component or cluster that student i is presumed to belong to. Each of the k components consists of a vector of Poisson rate parameters, $\boldsymbol{\lambda_k} = [\lambda_{k1}, \ldots, \lambda_{kd}, \ldots, \lambda_{k6}]$, where d from $\{1, \ldots, 6\}$ represents a specific day of the week. For example, one component could have very low values for all the λ_{kd}'s, for students with low daily activity, and another component could have high values for all the λ_{kd}'s, for students with high daily activity.

When fitting our mixture model to data, we take a Bayesian approach [13] and use Gamma prior distributions for the rate parameters λ_{kd}. The primary reason for doing this is to encourage the model to avoid degenerate solutions with a small component that has one or more rate parameters λ_{kd} at or near a value of 0. This can produce a high-likelihood solution but one that is not useful. In our experimental results later in the paper we used hyperparameters of $\alpha = 1.1$ and $\beta = 0.1$ for the Gamma distribution. These hyperparameter choices have the effect of making the Gamma prior behave like a step function, putting zero probability mass at $\lambda_{kd} = 0, k = 1, \ldots, K, d = 1, \ldots, 6$, and a relatively flat uninformative prior distribution over positive rate parameter values. Figure 3 depicts a graphical model representation of the Poisson mixture model with a Gamma prior on the $\boldsymbol{\lambda}$ parameters for each component.

The likelihood for the data $\boldsymbol{y_i}$ for each student i under this mixture model can be written as

$$p(\boldsymbol{y_i}|\boldsymbol{\lambda}) = \sum_{k=1}^{K} p(\boldsymbol{y_i}|z_i = k, \boldsymbol{\lambda_k})p(z_i = k) \qquad (1)$$

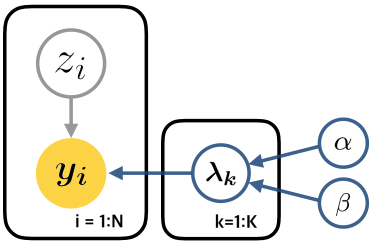

Figure 3: Graphical representation of the Poisson mixture model with Gamma prior. $\boldsymbol{y_i}$ and $\boldsymbol{\lambda_k}$ are 6 dimensional vectors. N is the number of students, and K is the number of mixture components.

where $p(z_i = k)$ is the marginal mixing weight for each component, and each component distribution can be written as

$$p(\boldsymbol{y_i}|z_i = k, \boldsymbol{\lambda_k}) = \prod_{d=1}^{D} p(y_{kd}|\lambda_{kd}, z_i = k) \quad (2)$$

assuming conditional independence of the daily counts y_{kd} given component k. λ_{kd} is the rate for day d for component k and each distribution $p(y_{kd}|\lambda_{kd}, z_i = k)$ is a Poisson distribution. The prior distribution is defined as a product over independent Gamma priors, one for each λ_{kd}, each with parameters $\alpha = 1.1$ and $\beta = 0.1$.

3.3 Learning Parameters with the EM Algorithm

To estimate the parameters $\boldsymbol{\lambda_k}$ of our model we use the Expectation-Maximization (EM) algorithm, an iterative algorithm that is widely used in fitting mixture models to data [8, 18]. More specifically, we use the EM algorithm to maximize the product of the data likelihood times the prior (both defined above). This results in both (a) maximum a posteriori (MAP) parameter estimates for the Poisson components in the model, and (b) membership weights w_{ik} that reflect the probability (under the fitted model) that each student i belongs to component (or cluster) k.

Each iteration of the EM algorithm consists of two steps, the E (expectation) step and the M (maximization) step. In the E-step, conditioned on some fixed (current) values of the parameters, the probability of membership w_{ik} is computed for each component $k = 1, \ldots, K$, for each student $i = 1, \ldots, N$.

$$\begin{aligned} w_{ik} &= p(z_i = k|\boldsymbol{y_i}, \boldsymbol{\lambda}, \alpha, \beta) \\ &\propto p(\boldsymbol{y_i}, z_i = k, \boldsymbol{\lambda_k}|\alpha, \beta) \\ &\propto p(\boldsymbol{y_i}|z_i = k, \boldsymbol{\lambda_k})p(\boldsymbol{\lambda_k}|\alpha, \beta)p(z_i = k) \end{aligned} \quad (3)$$

These membership weights are important in our later analyses, since they provide information of how likely it is that each data point i (in our case, student i) was generated by component k. In the M-step, conditioned on the set of membership probabilities w_{ik}, a point estimate of each parameter is estimated via MAP estimation.

$$\hat{\boldsymbol{\lambda}}_{\boldsymbol{k}} = \frac{\sum_i w_{ik}(\boldsymbol{y_i} + \alpha - 1)}{\sum_i w_{ik}(1 + \beta)} \quad (4)$$

$$\hat{p}(z_i = k) = \frac{\sum_i^N w_{ik}}{N} \quad (5)$$

These MAP parameter estimates provide the input for the next E-step, and thus, the cycle of E and M-steps continue iteratively.

The EM algorithm as a whole consists of randomly initializing the parameters of the model, followed by repeated computation of pairs of E and M steps, until the log-likelihood is judged to have converged (i.e., when the improvement in log-likelihood from one iteration to the next is less than some small value ϵ, or when the average membership probability value is not changing significantly from one iteration to the next).

Python code for this EM algorithm is available online at `https://github.com/jihyunp/student_poisson_mixture`.

4. DATA SETS

Two data sets from the same undergraduate online course were used in this study: one from summer 2016, and the other from summer 2017. Both summer courses were 5 weeks long. While each class was taught by two different instructors, the class content, such as the lecture videos, resources, and assignments, were the same. In both classes, students were assigned 5 video lectures every week and each lecture video had a corresponding quiz. The instructors encouraged students to watch one lecture video and complete the corresponding quiz each day, from Monday through Friday.

Although students were encouraged to follow this schedule, the actual deadline for watching the 5 lecture videos and completing the quizzes was on Fridays at midnight. While this structure gave students freedom to watch the lecture videos when they wanted, this flexibility also allowed them to procrastinate.

Most of the students' activities were recorded through the Canvas Learning Management System (LMS). These activities included downloading course content, watching lecture videos, taking online quizzes, submitting assignments, etc. Every time a student clicked on a URL within the Canvas system, the click event was logged with the student ID, URL, and time-stamp. The clickstream data was processed so that it only focused on the activities of daily tasks, resulting in *daily activity counts*, as mentioned in the previous section (Figure 1 and Figure 2). Only one event per task was counted and thus the sum of the matrix for each student was 25 or less (for 5 video lectures × 5 weeks). We chose to count only the first attempt (first click event) for each task.

In addition to the clickstream data, student demographic data was available through the university's institutional research office. It included both demographic information (gender, ethnicity, first generation status, low income status, and full-time status) and prior academic achievement (total SAT[1] score). Some students did not agree to provide this demographic data, although most did. For this reason, our later analyses based on demographic information are based on the subset of students who agreed to share this information.

[1]A standardized test widely used for college admissions in the United States.

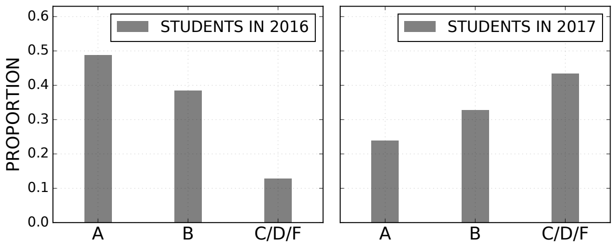

Figure 4: Grade distributions of students in 2016 class (left) and in 2017 class (right). Two classes show very different grade distributions. Almost half of the students received an A for the class in 2016, whereas more students got lower grades in 2017.

Although both classes used the same materials and implemented the same deadlines, there were some notable differences in how the the click events were recorded, as well as how each instructor structured the course. We describe these differences in the following sections.

4.1 Class in 2016

Online lectures and daily quizzes were offered outside the Canvas LMS for this class. Each lecture video was embedded on a separate web page on the server that we had access to, and the links to the web pages were provided via the Canvas weekly module.

Logs for the daily quiz attempts were not accessible, so instead we used the first "video clicks," which are from the logs of HTTP GET requests of the video embedded web pages. For each student, we matched the IP addresses of the video logs (from the server) with the IPs recorded on the Canvas LMS.

After removing 4 students with very low activity (0 or 1 video clicks in total) there were 172 students with activity counts available for analysis. More than 90% of the students received a passing grade, and half of the students received an A (Figure 4). Completing the daily tasks (watching videos and solving quizzes) counted as 15% towards the overall grade for each student.

4.2 Class in 2017

The video click logs for this class were not available since the videos were uploaded on a third-party server. However, the daily quizzes that students took after watching the lecture videos were recorded through the Canvas system, and we were able to obtain students' quiz submission time-stamps via the corresponding clickstream data. Therefore, for this class we focused on the first clicks for daily "quizzes." Note that this is different from the 2016 class data, which used the first clicks for each video-watching event.

There were 145 students in the class—we used data for 140 students after dropping 5 students with very low activity (as with the 2016 class). As previously noted, a different instructor taught the class in 2017 than in 2016. The instructor for the 2017 class changed the contribution to 8% of the total grade for watching and completing the lecture

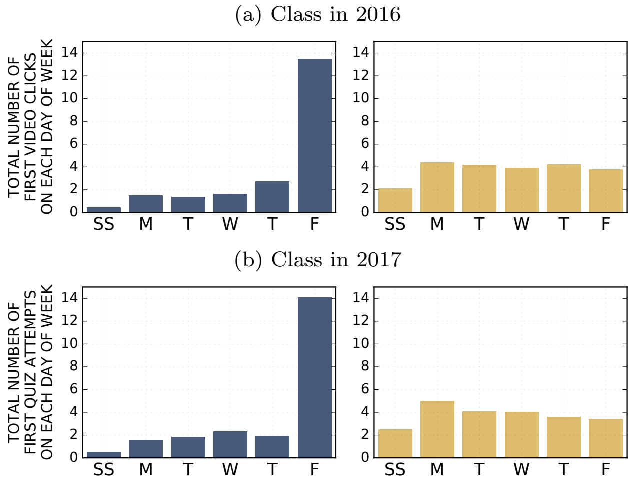

Figure 5: Poisson mixture component means ($\boldsymbol{\lambda}_k$'s) from modeling aggregated daily task counts ($\boldsymbol{y}_i$) for the class in 2016 (upper) and 2017 (lower).

videos, significantly less than in the 2016 class (15%). The grade distribution of the 2017 class in Figure 4 is also significantly different to that in 2016—there are significantly fewer students who received A's or B's in 2017 compared to the 2016 class.

5. PROCRASTINATION AS A MIXTURE COMPONENT

Below, we present and discuss the results of fitting a two-component ($K = 2$) Poisson mixture model to the *aggregated daily task counts* for the two classes described in section 4. We also explored models with more components, $K = 3, 4, \ldots$, but found that the $K = 2$ model broadly captured the primary modes of student behavior and that higher values of K tended to split the two main modes into further subgroups without providing any significant additional insight.

Figure 5 shows the expected number of counts per group, i.e., the rate parameters, $\boldsymbol{\lambda}_k$'s. The two group-dependent rate patterns across the days of the week, for both 2016 and 2017, show two very distinct behavioral patterns. One of the mixture components has a very high rate on Friday and low rates on the other days of the week. The other component has low and relatively flat rates from Monday to Friday. Considering the fact that the deadline for daily tasks in these courses is on Fridays, these two patterns clearly reflect two different types of student behaviors: *procrastination* and *non-procrastination*.

5.1 Characteristics of the Two Behavioral Groups

We can threshold the membership weights at 0.5 to classify each student $i = 1, \ldots, N$ into one of the two groups, i.e., if $w_{i1} > 0.5$ then student i is assigned to the *procrastination* group (where $k = 1$ corresponds to the *procrastination* group). About 36-37% of the students were assigned to the *procrastination* group in each of the two years.

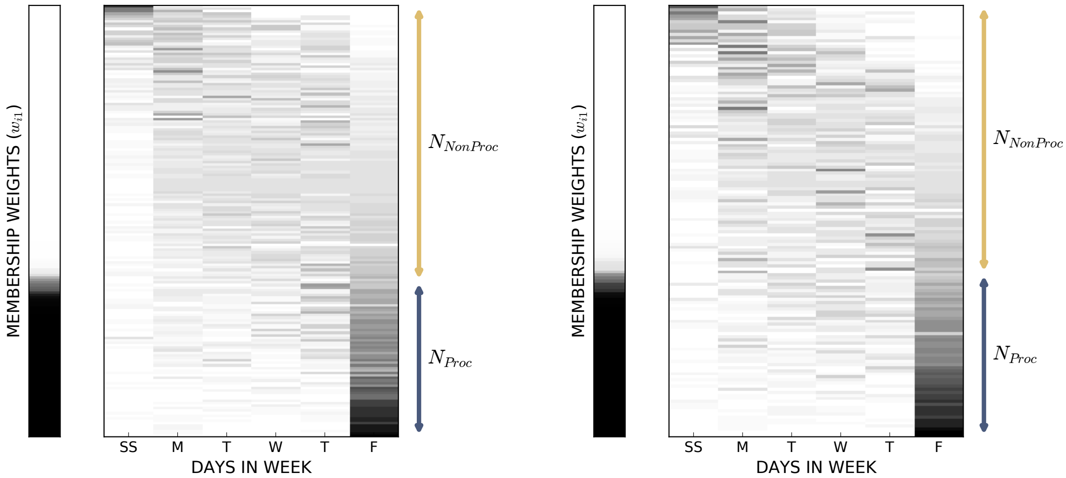

Figure 6: Aggregated daily task counts shown along with the membership weights. Each row represents a student, and the students are sorted by the membership weight w_{i1}. The left figure is for the class in 2016, and the right figure is for the class in 2017.

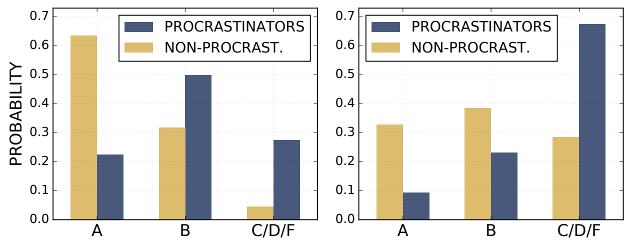

Figure 7: Probability of receiving each grade given that the student is in the *procrastination* group or in the *non-procrastination* group in 2016 (left) and in 2017 (right).

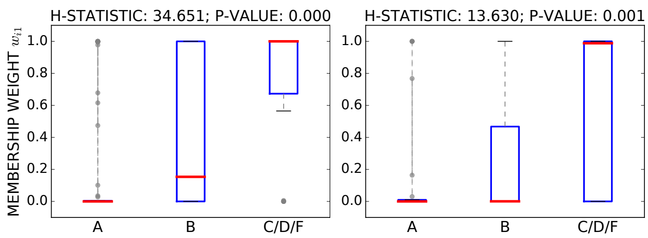

Figure 8: Distribution of w_{i1} in different grade groups of class in 2016 (left) and in 2017 (right). H-statistic comes from a Kruskal-Wallis test. (w_{i1}: membership weight on the procrastinating group)

The two plots in Figure 6 illustrate the students' week-aggregated activities along with the students' membership weights. Each row in each plot represents a student and the wider matrix plot shows the aggregated daily counts, sorted by their membership weight w_{i1}. The values in the matrix range from 0 to 25 and a darker color means that there are more task activities on that day of the week. The two plots from different years look almost identical and they clearly show the two types of behavior. The students (rows) at the bottom of each plot have more counts (darker colors) on Fridays and belong to the *procrastination* group. There is also a small group of students at the top of both plots who tend to be more active over the weekend. The size of this group of students is relatively small and their behavior pattern is effectively that of *non-procrastinators* since they are the "early birds" who check out the lecture videos or the quizzes early in the week.

The membership weights are shown on the narrower bar plot (left of each year's plot), where a darker color represents a higher membership weight of belonging to the procrastination group (with a weight close to 1). We can observe that there is a relatively small amount of grey area in the bar plot (for both years), which means that the majority of the students have a very high probability of being assigned to one group or the other.

5.2 Association between Behaviors and Grades

We can further analyze the relationship between the two different behavioral groups and the grades. We show the grade distribution in each group in Figure 7. Results from the two classes are shown side by side. It is obvious from the figure that the *non-procrastinators* tend to get significantly more A grades than the *procrastinators*, whereas the *procrastinators* get more C, D, and F's. Even though the overall grade distributions were quite different in the two classes (see Figure 4), we find a strong correlation between the behavioral groups and course outcomes. In both classes, the non-procrastinating students are about three times more likely to get an A grade than the procrastinating students. These probabilities were significant at the 0.01 level using a chi-squared test.

We can further analyze the relationship between procrastination behavior and grade outcomes by grouping students by their grade (rather than by the behavioral group) and looking at the patterns of behavior for each grade group.

As we saw in Figure 4, a majority of the students got a passing grade in 2016. The number of students who received A,

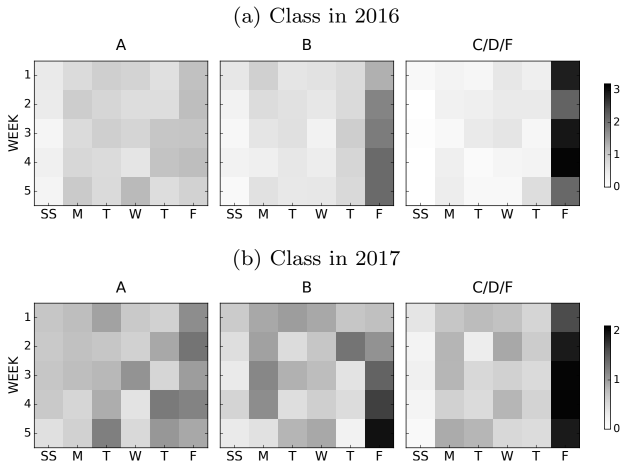

Figure 9: The number of task counts per day, for each of the 5 weeks, averaged over the students in each grade group. Left: students who received A's, middle: students who received B's, right: students who received a C, D, or F.

B, and the others (C, D, or F) were 84, 66, and 22, respectively. The left boxplot in Figure 8 informs that "A students" have very low membership weight (w_{i1}) values, but the "C, D, F students" have very high membership weight values. This can be interpreted as saying that the students who received lower grades (C, D, or F) have higher probabilities of being *procrastinators*.

We can see the similar result for the class in 2017 from the right side of the plot in Figure 8. There were 27, 37, 49 students in each of the grade groups (there were 27 students whose grade information was unavailable). The broader distribution of weights in the C, D, F group may be due to the fact that there were many more students with lower grades than higher grades in this year of the course.

The association between behavior and grade outcome is also clearly visible in the raw data, i.e., the task activity counts, for both years. Figure 9 clearly illustrates the behavior patterns for students with different grades. We can see a very dark color on Fridays on the matrices on the right side (students who received C, D, or F grades), and more evenly distributed colors on the left matrix plots, which shows the activities of students who received A grades.

6. REGULARITY OF PROCRASTINATION

In the previous section we showed that Poisson mixture modeling can help to unveil two latent types of students: *procrastinators* and *non-procrastinators*. Because these results are based on modeling *aggregated* daily activity counts across multiple weeks, they do not shed light on how students might change their procrastinating behavior over time during different stages of a course. For example, a student who is generally a procrastinator might only procrastinate every other week, while a non-procrastinator might postpone studying during some week. To gain insights into these nuances, we investigate the *regularity of procrastination* in this section.

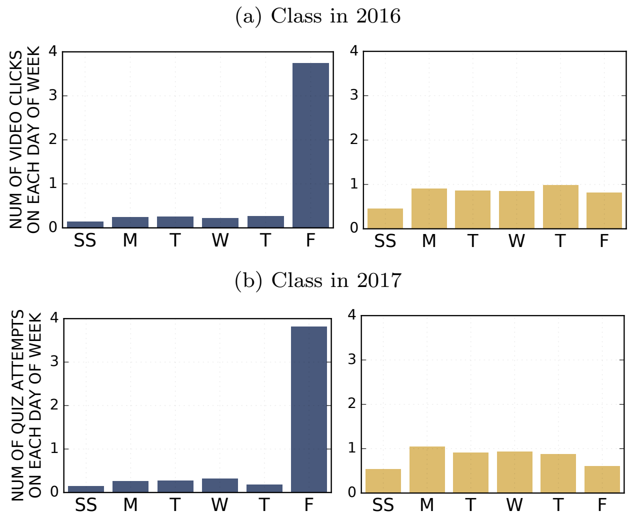

Figure 10: Poisson mixture component means (λ_k's) from modeling individual week of daily task counts (y_{ij}) for the class in 2016 (upper) and 2017 (lower). The number of first clicks on any lecture video in 2016, and the number of first attempts on any quiz in 2017, are modeled.

6.1 Regularity across Weeks

We focus here on inter-week regularity, which is defined as the extent that students repeat their behavior across different weeks. We use the same Poisson mixture modeling methodology described earlier in the paper except that we model each *individual week* of daily activity counts for each student rather than aggregating across weeks. The resulting mixture components are similar to the aggregated case in that there are two distinct weekly behaviors, *procrastination* and *non-procrastination* (see Figure 10). Each week of a student's behavior is modeled as being generated by one of the two components in the model, and we can estimate the membership weight of belonging to the *procrastination* group (or component) for each week for each student, i.e., w_{ij1} for student $i = 1, \ldots, N$ and week $j = 1, \ldots, M$ where $M = 5$ is the number of weeks.[2]

To quantify student i's regularity, we use the standard deviation of the *procrastination* weights across weeks:

$$SD_i = \left(\frac{1}{M-1} \sum_{j=1}^{M} (w_{ij1} - \overline{w}_{i \cdot 1})^2 \right)^{1/2} \tag{6}$$

where w_{ij1} and $\overline{w}_{i \cdot 1}$ represent the student-week membership weights for the *procrastination* component in week j, and the average of those weights across the M weeks, respectively.

[2]We could also use the non-aggregated *weekly* daily activity counts for the earlier group analyses in Section 5. Instead of the membership weights w_{i1} for student i, the mean value of the M membership weights ($\overline{w}_{i \cdot 1}$) could be used for thresholding. This would allow us to use the same analyses in Section 5 and 6 by fitting a single mixture model. We investigated this and found the results were almost identical to those reported in the paper. Given this, for the investigation of regularity we used weekly activity counts to see changes in weekly behavior, and for overall clustering (Section 5) we used total aggregate counts for ease of interpretation.

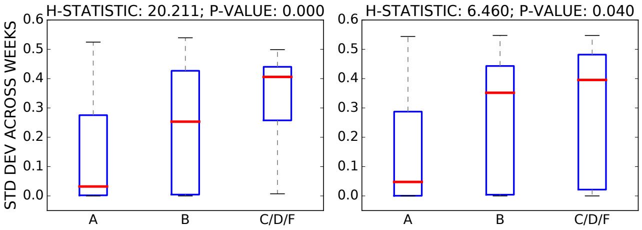

Figure 11: Distribution of SD_i in different grade groups of class in 2016 (left) and in 2017 (right). H-statistic comes from a Kruskal-Wallis test. (SD_i: inter-week standard deviation of w_{ij1}, membership weight on the *procrastination* group in week j)

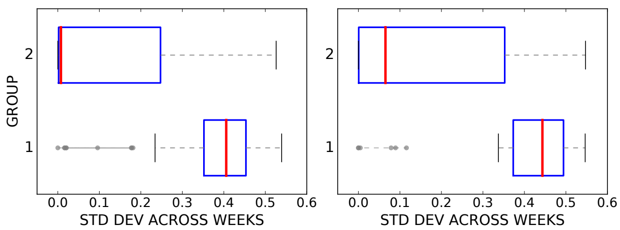

Figure 12: Distribution of SD_i in two behavioral groups (see Figure 6) of class in 2016 (left) and in 2017 (right). Within each subgraph, the *procrastination* group is indexed as 1, while the *non-procrastination* group is indexed as 2. (SD_i: inter-week standard deviation of w_{ij1}, membership weight on procrastination group in week j)

By definition, a higher value for SD_i signifies more volatile behavioral patterns.

In light of prior research, regularity is strongly correlated to performance [3]. We plot the distribution of SD_i's within three grade groups in Figure 11. Consistent with prior findings, students with better grades in general have lower levels of SD_i, hence are more regular learners. More formally, we perform a Kruskal-Wallis test within each class, with results reported above the graph. In both years, the three groups have significantly different SD_i distributions.

6.2 Incorporating Regularity and Procrastination

In previous sections, w_{i1} and SD_i capture different dimensions of procrastinating behavior, and their interaction is worth discussing further. For one thing, procrastinators and non-procrastinators may have different levels of regularity. We compare the distribution of SD_i within each behavioral group (assigned identically as in Figure 6) and plot the results in Figure 12. Common to both classes, *procrastinators* are centered around 0.4, while *non-procrastinators* on average have very small values below 0.1. We also calculate Pearson's correlation coefficient between w_{i1} (continuous membership weights before hard group assignments, as defined in Section 5) and SD_i, resulting in values of 0.675 for 2016 and 0.590 for 2017, both statistically significant at the 0.001 level. From these results, we can conclude

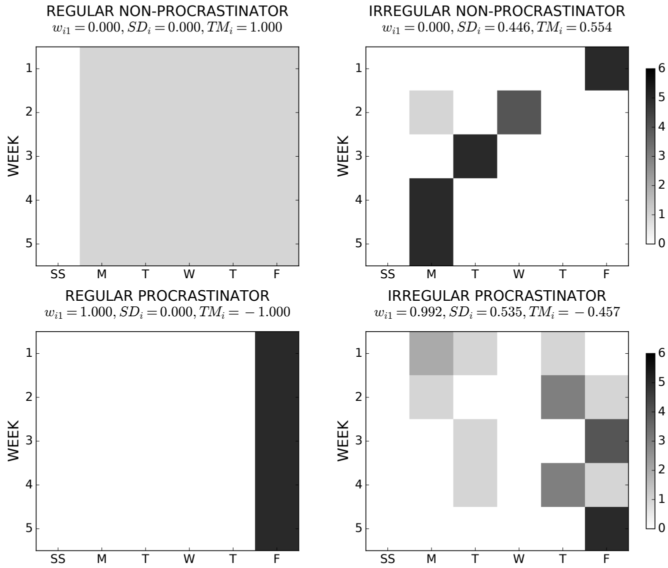

Figure 13: Number of daily activity counts for four prototypical students in the 2016 class. (w_{i1}: aggregated membership weight on *procrastination* group; SD_i: inter-week standard deviation of w_{ij1}, membership weight on *procrastination* group in week j; TM_i: Time Management Score)

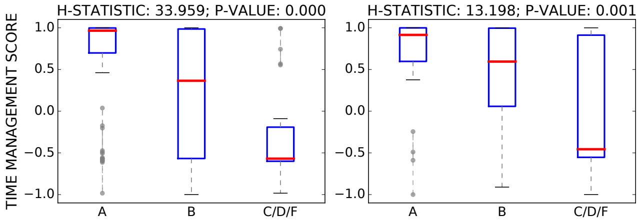

Figure 14: Distribution of Time Management Score (TM_i) in different grade groups of class in 2016 (left) and in 2017 (right). H-statistic comes from a Kruskal-Wallis test.

that non-procrastinating students are also more likely to stay consistent throughout the course, while procrastinators jump between spacing out their studies and postponing everything until the last day. On the other hand, the relationship between regularity and academic performance may substantially vary depending on how much a student is a procrastinator. Procrastinators who put off studying as a habit (with high regularity) may be more at-risk than those who occasionally jump to a spaced-out pattern, while this is the opposite for non-procrastinating students. To incorporate this asymmetry, we attempt to define a single index built upon w_{i1} and SD_i. As these two measures are both conceptually related to time management abilities, we name the index to be the Time Management Score (TM_i). To reflect their interaction, we multiply variations of w_{i1} and SD_i for each student i. Since w_{i1} and SD_i are both negatively correlated with outcome, we use the negative of their values in the index to allow for more natural interpretation. Moreover, because the score should be weighted in opposite directions depending on the student's behavioral group, we made variations to w_{i1} so that the most procrastinating student with the same degree of regularity would have the smallest score. Taking all of the above into account, we define TM_i as follows:

$$\begin{aligned} TM_i &= (1 - w_{i1})(1 - SD_i) + [-w_{i1}(1 - SD_i)] \\ &= (1 - 2w_{i1})(1 - SD_i) \end{aligned} \tag{7}$$

To evaluate the validity of this index, we examine whether its properties are aligned with theoretical assumptions. As discussed above, from the perspective of academic success, higher regularity is a negative behavioral feature for procrastinators but is a positive feature for non-procrastinators. In this context, it is natural to investigate how regularity and procrastination affects the value of TM_i, and how this value relates to desirable and undesirable outcomes.

From Equation (7), we know that the value of TM_i is positive or negative depending on whether w_{i1} is greater than 0.5 or not. Thus, $w_{i1} = 0.5$ is the watershed of whether SD_i positively or negatively contributes to TM_i. Given that our threshold for hard group assignment in Section 5 is also 0.5, the interpretation is straightforward: higher regularity leads to higher TM_i within the *procrastination* group, and it is more so for "purer" procrastinators; the opposite story can be told within the *non-procrastination* group.

For an intuitive examination, we choose four prototypical students with different levels of procrastination and regularity from the classs in 2016, and plot the daily counts of their first video clicks in Figure 13, along with their w_{i1}, SD_i and TM_i. As we would expect, non-procrastination with high regularity (upper-left), the most desirable pattern, has $TM_i = 1$, the maximum value possible in our context. By contrast, the regular procrastinator (lower-left) gets the minimum value of $TM_i = -1$. The remaining two students with similarly low regularity have TM_i values between the two extremes, but are respectively closer to the one that belongs to the same behavioral group. In a word, these visual patterns further validate the construction of TM_i, which more precisely measures the degree of procrastination by incorporating regularity information.

To determine if TM_i captures the desirability of certain procrastinating patterns, we probe into the relationship between this index and course outcomes. Similar to what we did earlier with w_{i1} and SD_i individually, we plot the distribution of TM_i within three grade groups. As shown in Figure 14, there exists a positive relationship between TM_i and performance, which is statistically significant under a Kruskal-Wallis test. The TM_i score incorporates two measures (w_{i1} and SD_i) and amplifies the information that is potentially predictive of performance, providing a more nuanced view of procrastination.

7. RELATIONSHIP WITH STUDENT BACKGROUND

Having explored the fine-grained differences in students' procrastinating behaviors and their relationship with outcomes, we want to further examine if these variations can be discriminated by students' background characteristics. The goal of this analysis is to understand whether there exists a potential risk factor among underrepresented students, or if instead, the behavioral differences we observe are more individual-level in nature. We also sought to explore whether prior academic achievement could explain differences in procrastinating behaviors.

Table 1: Relationship between demographic variables and procrastination/regularity measures for the 2016 class

(a) Behavioral group assignment (binary)

Demographics	N	Test	p-value
$FirstGen$	144	χ^2-test	0.566
$LowInc$	151		0.672
SAT	147	K-W test	0.238

(b) SD and TM (continuous)

Demographics	N	Test	SD p-val	TM p-val
$FirstGen$	144	K-W test	0.884	0.954
$LowInc$	151		0.175	0.294
SAT	147	Pearson's r	0.118	0.363

Table 2: Relationship between demographic variables and procrastination/regularity measures for the 2017 class

(a) Behavioral group assignment (binary)

Demographics	N	Test	p-value
$FirstGen$	120	χ^2-test	0.218
$LowInc$	128		0.955
SAT	125	K-W test	0.802

(b) SD and TM (continuous)

Demographics	N	Test	SD p-val	TM p-val
$FirstGen$	120	K-W test	0.136	0.897
$LowInc$	128		0.754	0.973
SAT	125	Pearson's r	0.505	0.820

From a comprehensive list of demographic variables, we choose three that are of general interest in education research: *Low Income Status*, *First Generation* and *Total SAT Score*. The first two binary variables represent a student's social-economic status, and the last continuous variable is a proxy for prior academic achievement.

We separately test the relationships between these three variables and three measures of procrastination and regularity in previous sections: behavioral group assignment (as in Section 5.1), SD and TM (as in Section 6). The specific statistical tests we use and their results are reported in Table 1 for the class in 2016, and Table 2 for the class in 2017. Because the demographic information contains missing values, we only include students who have relevant information in each of the tests (the number of students, N, is reported in the tables).

The results show that for both classes none of these demographic variables have any significant relationship with procrastination and/or regularity. This suggests that failures in time management may arise more from students' inherent factors than specific background characteristics, and that effective instructional interventions are less likely to be hampered by students' underrepresented backgrounds. However, due to the limited class sizes, this inference still needs to be further explored at scale.

8. CONCLUSIONS

In this paper, we introduce a data-driven methodology for characterizing student procrastination in online courses. Based on Poisson mixture modeling, the proposed approach can be applied to courses where tasks with clear deadlines are regularly assigned and students' timestamped activites related to those tasks are recorded. In our experiments with two undergraduate online classes, this method identifies two distinct patterns in students' weekly planning behavior, which can be further utilized to measure procrastination. This measure is found to be strongly correlated with course outcomes for both classes. In addition, our proposed Time Management Score (TM) is able to quantify students' overall time management skills by combining overall degree of procrastination with the regularity of the behavior. Interestingly, while TM is a strong predictor of course outcomes, it is not significantly related to students' demographics or prior academic achievement. These results suggest that, as a whole, procrastination behaviors seem to be more of an inherent characteristic.

These types of clickstream data and analyses allow for rich complements to other types of educational research. For example, the proposed behavioral measures of time management can be combined with survey data to examine how accurate students' perceptions of their skills are, and to identify students who might be especially prone to benefit from support. From the practical perspective, these data-driven approaches can be incorporated into learning management systems and work in real time. This would potentially facilitate automated assessment and intervention regarding time management skills.

There are also a number of potentially useful extensions to the methodological approach proposed here. For example, the mixture components in the two classes that we analyzed are straightforward to interpret with regard to procrastination, but this might not be the case for different course designs and structures. In these broader scenarios, it may be useful to incorporate informative Gamma prior distributions into the mixture model, with, for instance, three prior components for procrastination behavior, non-procrastination behavior, and mixed behavior respectively.

9. ACKNOWLEDGMENTS

This paper is based upon work supported by the National Science Foundation under Grants Number 1535300 (for all authors) and 1320527 (for PS). The authors would like to thank the Durendal Van Huynh, Kelsey Hollis Layos, and Qiujie Li for their assistance in acquiring the data used in this paper and giving helpful advice. We would also like to thank the course instructors, Dr. Jeremy Eaton and Dr. Anna Kwa, for their support and intellectual assistance.

10. REFERENCES

[1] A. R. Anaya and J. G. Boticario. A data mining approach to reveal representative collaboration indicators in open collaboration crameworks. In *Proceedings of the 2nd International Conference on Educational Data Mining*, pages 210–219, Cordoba, Spain, 2009.

[2] E. P. Bettinger, L. Fox, S. Loeb, and E. S. Taylor. Virtual classrooms: How online college courses affect student success. *American Economic Review*, 107(9):2855–75, 2017.

[3] M. S. Boroujeni, K. Sharma, Ł. Kidziński, L. Lucignano, and P. Dillenbourg. How to quantify student's regularity? In *Proceedings of the 11th European Conference on Technology Enhanced Learning (EC-TEL 2016)*, pages 277–291, Lyon, France, sep 2016. Springer, Cham.

[4] T. Brijs, D. Karlis, G. Swinnen, K. Vanhoof, G. Wets, and P. Manchanda. A multivariate Poisson mixture model for marketing applications. *Statistica Neerlandica*, 58(3):322–348, 2004.

[5] J. Broadbent and W. Poon. Self-regulated learning strategies & academic achievement in online higher education learning environments: A systematic review. *The Internet and Higher Education*, 27:1–13, 2015.

[6] J. M. Calabrese, J. L. Brunner, and R. S. Ostfeld. Partitioning the aggregation of parasites on hosts into intrinsic and extrinsic components via an extended Poisson-gamma mixture model. *PLoS ONE*, 6(12):1–9, 2011.

[7] X. Chen and C. Carroll. Fgs in post-secondary education: A look at their college transcripts (nces 2005-171). us department of education. *National Center for Education Statistics. Washington, DC: US Government Printing Office. Retrieved from http://nces. ed. gov/pubs2005/2005171. pdf*, 2005.

[8] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B (Methodological)*, 39(1):1–38, 1977.

[9] T. Dvorak and M. Jia. Online work habits and academic performance. *Journal of Learning Analytics*, 3(3):318–330, 2016.

[10] G. C. Elvers, D. J. Polzella, and K. Graetz. Procrastination in online courses: Performance and attitudinal differences. *Teaching of Psychology*, 30(2):159–162, 2003.

[11] W. Feng, Y. Liu, J. Wu, K. P. Nephew, T. H. M. Huang, and L. Li. A Poisson mixture model to identify changes in RNA polymerase II binding quantity using high-throughput sequencing technology. *BMC Genomics*, 9(Suppl 2):S23, 2008.

[12] C. Fraley and A. E. Raftery. Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association*, 97(458):611–631, 2002.

[13] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian data analysis.* CRC press, 2013.

[14] S. L. Hotle. *Applications of clickstream information in estimating online user behavior.* PhD thesis, Georgia Institute of Technology, 2015.

[15] R. R. Jayasekare, R. Gill, and K. Lee. Modeling discrete stock price changes using a mixture of Poisson distributions. *Journal of the Korean Statistical Society*, 45(3):409–421, 2016.

[16] A. M. Kazerouni, S. H. Edwards, and C. A. Shaffer. Quantifying incremental development practices and their relationship to procrastination. In *Proceedings of the 2017 ACM Conference on International Computing Education Research*, ICER '17, pages 191–199. ACM, 2017.

[17] J. Li and H. Zha. Two-way Poisson mixture models for simultaneous document classification and word clustering. *Computational Statistics & Data Analysis*, 50(1):163–180, 2006.

[18] G. McLachlan and T. Krishnan. *The EM Algorithm and Extensions*, volume 382. John Wiley & Sons, 2007.

[19] P. D. McNicholas. Model-based clustering. *Journal of Classification*, 33(3):331–373, 2016.

[20] B. L. L. Ng, W. C. Liu, and J. C. K. Wang. Student motivation and learning in mathematics and science: A cluster analysis. *International Journal of Science and Mathematics Education*, 14(7):1359–1376, 2016.

[21] J. Park, K. Denaro, F. Rodriguez, P. Smyth, and M. Warschauer. Detecting changes in student behavior from clickstream data. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference on - LAK '17*, pages 21–30, Vancouver, BC, Canada, 2017. ACM Press.

[22] P. R. Pintrich. A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review*, 16(4):385–407, 2004.

[23] E. L. Snow, G. T. Jackson, and D. S. McNamara. Emergent behaviors in computer-based learning environments: Computational signals of catching up. *Computers in Human Behavior*, 41:62–70, Dec 2014.

[24] P. Steel and J. Ferrari. Sex, education and procrastination: an epidemiological study of procrastinators' characteristics from a global sample. *European Journal of Personality*, 27(1):51–58, 2013.

[25] K. M. Styck. Best practices for supporting upward economic and social mobility for first-generation college students. *The School Psychologist*, 72(2):50–57, 2018.

[26] K. Tóth, S. Greiff, C. Kalergi, and S. Wüstenberg. Discovering students' complex problem solving strategies in educational assessment. *Proceedings of the 7th International Conference on Educational Data Mining (EDM)*, pages 225–228, 2014.

[27] J. Whitehill, J. Williams, G. Lopez, C. Coleman, and J. Reich. Beyond prediction: first steps toward automatic intervention in MOOC student stopout. In *Proceedings of the 8nd International Conference on Educational Data Mining*, pages 222–230, Madrid, Spain, 2015.

[28] J. Xie, A. Essa, S. Mojarad, R. S. Baker, K. Shubeck, and X. Hu. Student learning strategies and behaviors to predict success in an online adaptive mathematics tutoring system. In *Proceedings of the 10th International Conference on Educational Data Mining*, pages 460–465, Wuhan, China, 2017.

[29] J. W. You. Examining the effect of academic procrastination on achievement using LMS data in e-learning. *Journal of Educational Technology & Society*, 18(3):64, 2015.

[30] J. W. You. Identifying significant indicators using lms data to predict course achievement in online learning. *The Internet and Higher Education*, 29:23–30, 2016.

[31] B. J. Zimmerman. Self-regulated learning and academic achievement: An overview. *Educational psychologist*, 25(1):3–17, 1990.