Utilizing Learning Analytics to Map Students' Self-Reported Study Strategies to Click Behaviors in STEM Courses

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

Mariela Janet Rivas University of California, Irvine mrivas3@uci.edu Renzhe Yu University of California, Irvine renzhey@uci.edu

Mark Warschauer University of California, Irvine markw@uci.edu Jihyun Park University of California, Irvine jihyunp@ics.uci.edu

Brian K. Sato University of California, Irvine bsato@uci.edu

ABSTRACT

Informed by cognitive theories of learning, this work examined how students' self-reported study patterns (spacing vs. cramming) corresponded to their engagement with the Learning Management System (LMS) across two years in a large biology course. We specifically focused on how students accessed non-mandatory resources (lecture videos, lecture slides) and considered whether this pattern differed by underrepresented minority (URM) status. Overall, students who self-reported utilizing spacing strategies throughout the course had higher grades than students who reported cramming throughout the course. When examining LMS engagement, only a small percentage of students accessed the lecture videos and lecture slides. Applying a negative binomial regression model to daily counts of click activities, we also found that students who utilized spacing strategies accessed LMS resources more often but not earlier before major deadlines. Moreover, this finding was not different for underrepresented students. Our results provide some initial evidence showing how spacing behaviors correspond to accessing learning resources. However, given the lack of general engagement with LMS resources, our results underscore the value of encouraging students to utilize these resources when studying course material.

CCS CONCEPTS

 Applied computing → Learning management systems; Elearning;

KEYWORDS

Study Skills; Spacing Effect; Learning Analytics; Higher Education; STEM Education; Underrepresented Students

ACM Reference Format:

Fernando Rodriguez, Renzhe Yu, Jihyun Park, Mariela Janet Rivas, Mark Warschauer, and Brian K. Sato. 2019. Utilizing Learning Analytics to Map Students' Self-Reported Study Strategies to Click Behaviors in STEM Courses. In *The 9th International Learning Analytics & Knowledge Conference (LAK19)*,

LAK19, March 4-8, 2019, Tempe, AZ, USA

© 2019 Association for Computing Machinery.

ACM ISBN 978-1-4503-6256-6/19/03.

https://doi.org/10.1145/3303772.3303841

March 4-8, 2019, Tempe, AZ, USA. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3303772.3303841

1 INTRODUCTION

Learning analytics can offer valuable insights into students' study behaviors. By examining when and how often students engage within a Learning Management System (LMS), for example, we can better understand how these behaviors are related to learning outcomes, and use this understanding to improve instruction and course design features. This may be particularly beneficial for understanding student achievement in college STEM courses, which place high demands on students' learning. Moreover, national reports highlighting the disparity in success for underrepresented minorities underscore the need to improve student achievement in these fields [8].

Efforts to improve learning in STEM courses have traditionally focused on increasing active learning in the classroom, which entails increasing faculty and peer interaction [4]. However, another approach has been to utilize cognitive theories of learning to foster successful study behaviors. One such behavior is spacing, which refers to splitting up study sessions across multiple days, as opposed to 'cramming' one's studying into a single session [1, 2]. One of the reasons that spaced studying benefits learning is that it allows one to forget information [3, 6]. Generally, the more time in between study sessions (the more forgetting), the better because the act of trying to remember and review information helps strengthen and solidify that information into memory. While cramming may work in the very short term, it is detrimental for meaningful, longterm learning [10, 12]. Cramming can be particularly detrimental for courses that have cumulative exams, which require students to remember more information and retain it for a longer period, compared to non-cumulative exams that only cover a fraction of the course material at a time.

Recent work has found that students who self-report engaging in spacing strategies are more academically successful than students who report cramming behaviors [7, 11]. However, one of the major limitations of using survey data to measure students' spacing practices is that, while survey data can reveal important relationships with learning outcomes, it does not provide a clear mechanistic understanding of how spacing strategies correspond to actual behaviors.

One way to address this is to utilize clickstream data that record students' fine-grained interactions with the LMS, since it can offer a unique window into students' learning behaviors. By mapping

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

students' self-reported learning strategies to their behavior within the LMS, we may better understand how students with different learning profiles engage with course resources, especially those the instructor provides to help students review course concepts. The insights gained from combining both data sources can additionally inform instructors about the specific types of resources students are actually using to prepare for exams, and whether the timing of accessing these resources explain differences in students' learning outcomes.

Other studies on student behaviors and measures of success exist in the learning analytics space. For example, recent work has demonstrated significant correlations between student procrastination, the regularity of that procrastinating behavior, and learning outcomes [9]. The current study bridges self-report data with learning analytic data to provide a better understanding of how spacing and cramming behaviors correspond to engagement within a course's LMS.

1.1 Study aims and research questions

Informed by cognitive theories of learning, the aims of this study were to understand how students' self-reported spacing patterns (spacing vs. cramming) reflected engagement, especially in the context of pending course deadlines and exams. Our research questions are as follows:

- (1) Do self-reported measures of spacing and cramming reveal differences in grades?
- (2) Are self-reported study patterns and student demographics suggestive of differences in actual learning behavior within the LMS?
- (3) How do multiple types of course deadlines shape students learning behavior? Does it differ between self-reported spacers and crammers?

2 STUDY SETTING AND PARTICIPANTS

This study took place in a four-year institution in the western United States. Data was collected from students enrolled in a 10 week-long, face-to-face, lower division molecular biology course in 2016 and 2017. This second-year course, consisting primarily of sophomore students, was taught in a highly structured format, with roughly 50 percent of the course period dedicated to lectures and 50 percent to interactive peer group work. The course had three cumulative midterm exams and a cumulative final. Data from Year 1 spanned 74 days, and for Year 2, there were 73 days.

There were three lectures each week. Prior to each lecture, students were expected to complete the assigned textbook reading and take a brief online quiz. Additionally, there were weekly homework assignments students needed to complete online. The instructor also provided students with lecture videos and slides as a resource for reviewing the course content. All these resources and tasks were placed in the course's Canvas LMS.

This study was based on previously published data that examined the relationship between an intervention on study skills and course outcomes [11]. For this study, we only examined two of the six courses from the original study. Both courses were intervention courses, and both contained LMS data that has not been extensively analyzed or published. There were 224 registered students in Year



Figure 1: The average overall clicks across 10 weeks for Year 1. The error bars represent standard errors. Dashed red lines: exam days; dashed green lines: homework due dates; solid light blue lines: quiz due dates.

1 and 422 in Year 2. Because our study relied on survey data to assess students' study strategies, we only included students who completed these surveys. Year 1 consisted of 132 students (58.92% response rate) and Year 2 consisted of 327 students (77.48% response rate).

3 DATA

3.1 Clickstream data

We obtained students' clickstream data from both courses, which contained time-stamped records of every click a student made within the LMS. Aside from some minor differences in exam scheduling, the Canvas LMS sites were virtually identical between Years 1 and 2.

Figure 1 depicts the number of all clicks by day for Year 1. As one can observe, there was a steady stream of course activity with notable peaks on the days leading up to different deadlines. Our main analysis was restricted to clicks on two key resources: lecture videos and lecture slides. We focused on these resources because we wanted to obtain click measures that reflected self-directed studying. In this course, lecture videos and slides were provided to help students reference and review course material on their own. We did not have direct access to click information for the lecture videos themselves. Thus, our lecture video measure was obtained by observing the number of times a student visited the lecture video webpage within the Canvas LMS. The lecture slides were posted as individual links within the LMS. Unless otherwise specified, we use the general term of "course resources" to refer to these two types of resources in the remainder of this paper.

3.2 Student background

We obtained institutional records to examine students' underrepresented minority status. Students were categorized as having underrepresented minority status if their ethnic category was African American, Native American, or Latino/a. Mapping Students' Self-Reported Study Strategies to Click Behaviors

	Maintained	Stopped	Started	Maintained
	Cramming	Spacing	Spacing	Spacing
	n (%)	n (%)	n (%)	n (%)
Year 1	37	7	16	72
N = 132	(28.03%)	(5.30%)	(12.12%)	(54.54%)
Year 2	104	45	51	127
N = 327	(31.80%)	(13.76%)	(15.59%)	(38.83%)

 Table 1: The number and percentage of students who fell

 into the four study pattern categories for Years 1 and 2.

3.3 Learning outcomes

Students' learning outcomes were assessed by the final grade they received in the course. The scoring for these grades were as follows: 1 = F, 2 = D-, 3 = D, 4 = D+, 5 = C-, 6 = C, 7 = C+, 8 = B-, 9 = B, 10 = B+, 11 = A-, 12 = A, 13 = A+.

4 STUDY STRATEGY SURVEY MEASURE

We used a slight modification of a study strategies survey developed for college students [5, 7]. Our measure of students' study patterns (spacing vs. cramming) was obtained by combining two survey questions. The first question asked students, "Which of the following best describes your study patterns?," and they could select "I most often space out my study sessions over multiple days/weeks" or "I most often do my studying right before the test." The second question asked students to select the top 3 study strategies they used from a list. One of the selection options was "Absorbing lots of information the night before the test." Students were categorized as spacers if the only item they selected was "most often space out my study sessions over multiple days/weeks." Students were categorized as crammers if they selected either one or both of the cramming items.

Because students were surveyed at the beginning and the end of the course, we were able to categorize students into four study pattern categories: maintained cramming, stopped spacing, started spacing, maintained spacing. Students who reported cramming in both pre- and post-surveys were categorized as *maintained cramming*. Those who reported spacing in the pre-survey but cramming in the post-survey were categorized as *stopped spacing*. Students with the opposite pattern were categorized as *started spacing*. Finally, those who reported spacing in both pre- and post-surveys were categorized as *maintained spacing*.

Table 1 provides the number and percentage of students who fell into these categories. Given that we were interested in understanding differences among students with consistent study patterns, we focus on the students who reported maintaining spacing and maintaining cramming in our analyses of grade differences.

5 STATISTICAL METHODS

As the basis of modeling students' click activities on course resources, we calculated *daily activity count* – the number of clicks per day – for each student. We further constructed a daily activity count matrix $Y_{N\times T}$, where N is the number of students and T is the total number of days within the course. Each element of the matrix, y_{it} , represents the number of activities of student *i* on day *t*, where $1 \le i \le N$ and $1 \le t \le T$.



Figure 2: Distribution of daily activity counts on course resources, for Years 1 and 2.

To find out how student activities are related to individual and temporal characteristics, we modeled the daily activity counts using a negative binomial regression model due to the excess zeros in these counts. Figure 2 shows the distribution of the daily activity counts for the course resources in both years. The data has a high density of zeros since it was not mandatory to watch the videos or access the lecture slides. In the negative binomial distribution, there are two parameters: the mean *m* and the "dispersion parameter" *r*. In this setting, the variance $\sigma^2 = m + m^2/r$.

Letting μ_{it} be the mean of the negative binomial distribution for student *i* at time *t*, we modeled the logarithm of the mean $\log \mu_{it}$ as a linear function of covariates x_i and z_t as following:

$$\log \mu_{it} = \alpha + \beta x_i + \gamma z_t, \qquad (1)$$

where x_i is an array of student-level covariates and z_t represents time-specific covariates. α represents the intercept and β and γ represent the regression coefficients for the covariates.

We used glm.nb function from MASS package in R to fit the regression. The dispersion parameter r is estimated using a moment estimator after an initial fit of a Poisson regression model.

The likelihood for a single count y_{it} becomes

$$P(y_{it}|\mu_{it},r) = {y_{it} + r - 1 \choose y_{it}} p_{it}^{y_{it}} (1 - p_{it})^r,$$
(2)

where

$$p_{it} = \frac{\mu_{it}}{\mu_{it} + r}.$$
(3)

6 RESULTS

We first wanted to understand whether students' self-reported study patterns were related to their final grade in the course. We specifically compared the grades of students who maintained cramming and those who maintained spacing. As shown in Figure 3, students who reported maintaining spacing strategies had higher grades overall than students who reported maintaining cramming, which demonstrates the theoretical benefits of spacing (t(107) = -3.36, p < .01 for Year 1; t(229) = -4.04, p < .001 for Year 2).

To delve into the patterns of accessing course resources among different groups of students and at different time points in relation to course deadlines, we fitted a separate negative binomial regression, as described in Section 5, for each year. We included the following covariates:

• Self-reported study pattern (student-level): categorical variable as described in Section 4, with *maintained cramming* as the reference group

LAK19, March 4-8, 2019, Tempe, AZ, USA



Figure 3: Boxplot of course grades by study patterns for Years 1 and 2. The red dot and numeric values represent the means for each group.

(a) Year 1				
Covariate	Coef.	Std. Err.	Pr(> z)	
Stopped Spacing	0.8586	0.0572	0.0000	
Started Spacing	0.9486	0.0423	0.0000	
Maintained Spacing	0.6243	0.0308	0.0000	
URM	-0.2725	0.0322	0.0000	
Days from next homework due	0.0575	0.0059	0.0000	
Days from next exam	-0.1693	0.0029	0.0000	
Days from next quiz due	0.0276	0.0119	0.0210	
(b) Year 2				
Covariate	Coef.	Std. Err.	Pr(> z)	
Stopped Spacing	0.2675	0.0307	0.0000	
-				

Started Spacing	0.6031	0.0284	0.0000
Maintained Spacing	0.8573	0.0222	0.0000
URM	0.0418	0.0210	0.0462
Days from next homework due	0.0382	0.0044	0.0000
Days from next exam	-0.1933	0.0023	0.0000
Days from next quiz due	0.1460	0.0084	0.0000

Table 2: Summary of negative binomial regression results for Year 1 (top) and Year 2 (bottom). Estimated dispersion parameter (r) values were 0.09867 and 0.0688, respectively.

- Under represented minority (URM) status (student-level): binary variable
- Days from next homework due date (time-specific): days remaining before the next homework is due at the current time point
- Days from next exam (time-specific): days remaining before the next exam at the current time point
- Days from next quiz due date (time-specific): days remaining before the next quiz is due at the current time point

The overall results for both years are shown in Table 2, where intercepts are left out for brevity. In both course offerings, we were able to observe that students' self-reported study patterns were a strong indicator of actual learning behavior. Students who selfreported that they spaced at least once throughout the course had significantly positive coefficients, which means that they engaged more with the course resources than the students who maintained cramming.



Figure 4: Average resource clicks across 10 weeks for Years 1 and 2, grouped by maintaining spacing (blue line) versus maintaining cramming (gold line). The error bars represent standard errors. The dashed red lines denote exam days.

Figure 4 shows the daily resource click counts over the days, averaged by the students in two consistent study pattern categories. The blue lines (*maintained spacing*) are above the gold lines (*maintained cramming*) most of the time in both years. Our regression analysis validates the differences between the two self-reported study pattern groups.

URM status did not show consistent results, since its coefficients across two different years had opposite directions. In addition, URM was less significant in Year 2. Therefore, it is hard to draw a conclusion on whether URM students access course resources differently from non-URM students.

The three time-relevant covariates were incorporated to tease out students' behavioral responses to different types of course deadlines. Because exams were given the largest weights in the course design, we assumed that approaching exam dates would be the strongest predictor of students' click activities on course resources. From the bottom three rows of each panel in Table 2, we can see significantly negative coefficients of the exam covariate which confirmed this assumption. The other two types of deadlines, however, showed positive coefficients. One possible explanation is that students might not revisit course resources for lower-stakes course assignments, but given the complicated overlapping structures of different deadlines, it still deserves further scrutiny.

We then investigated whether students with different self-reported study patterns were driven differently by the deadlines. Within each class, we fitted the same negative binomial regression for maintained spacers and maintained crammers, but only included time-relevant covariates. Tables 3 and 4 present the results from Year 1 and Year 2, respectively. Overall, the directions of these coefficients were consistent with what was observed from the entire sample in Table 2, further suggesting that only exam deadlines were driving students to access the course resources. Mapping Students' Self-Reported Study Strategies to Click Behaviors

(), , , , , , , , , , , , , , , , , , ,			
Covariate	Coef.	Std. Err.	Pr(> z)
Days from next homework due	0.0601	0.0112	0.0000
Days from next exam	-0.1539	0.0056	0.0000
Days from next quiz due	-0.0897	0.0232	0.0001
(1)			

(a) Year 1, Maintained Cramming, r = 0.0728

Days from next quiz due	-0.0897	0.0252	0.0001
(b) Year 1, Maintained Spacing, $r = 0.103$			
Covariate	Coef.	Std. Err.	Pr(> z)
Days from next homework due	0.0569	0.0079	0.0000
Days from next exam	-0.1826	0.0039	0.0000
Days from next quiz due	0.0220	0.0161	0.1703

Table 3: Summary of negative binomial regression results for students who maintained spacing and who maintained cramming, Year 1

(a) Year 2, Maintained Cramming, r = 0.0556

3 0.0083	0.0000
4 0.0042	0.0000
0 0.0155	0.0000
	3 0.0083 4 0.0042 0 0.0155

(b) Year 2, Maintained Spacing, $r = 0.077$			
Covariate	Coef.	Std. Err.	Pr(> z)
Days from next homework due	0.0096	0.0067	0.1503
Days from next exam	-0.1799	0.0033	0.0000
Days from next quiz due	0.0751	0.0130	0.0000

Table 4: Summary of negative binomial regression results for students who maintained spacing and who maintained cramming, Year 2

7 CONCLUSIONS

Confirming learning theories, our work demonstrates the positive relationship between spaced study sessions and learning in a large biology course. We found that students who reported that they maintained spacing strategies throughout the course had higher final grades than students who reported maintaining cramming behaviors.

Using clickstream data, we were able to observe some specific behavioral mechanisms at play. When accessing course resources (lecture videos, lecture slides), students who maintained spacing had higher overall clicks than students who maintained cramming behaviors. However, spacing and cramming students showed similarly significant patterns of cramming for exams. This finding points to the need of providing more targeted instruction that encourages students to space out their study sessions, especially for courses that place high demands on students' learning.

In terms of how to support underrepresented learners, our results did not provide clear insights. We found inconsistent results between underrepresented status and click activities. This may be partly due to the fact that this was a second-year course for biology majors, and the underrepresented students enrolled in this course had already demonstrated their ability to succeed in biology. Future work could benefit from looking at underrepresented students' study patterns in freshman-level courses, where they may be most at risk for failing.

Our study had some limitations. Since we did not have a full sample of students in the course, our results may not provide a complete picture of student differences in spacing and cramming behaviors. In addition, our work is limited in that students' study behaviors (e.g., reviewing textbook chapters, notes, etc.) largely existed outside of the LMS. Therefore, we do not assume that the results obtained from this data represent the full spectrum of students' study behaviors, or that these click behaviors are indicative of success in the course. Instead, we show how students' self-reported study patterns correspond to differences in accessing resources when preparing for assignments and exams.

Finally, while lecture videos and slides are useful learning resources, they may be underutilized by students. Based on the results from this work, we recommend that instructors integrate these resources in the LMS in a way that encourages students to incorporate them into their regular study practices.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation (Grant No. 1535300). We would like to thank Preston Reed, Kelsey Hollis Layos, and Briandy Walden for their help acquiring project data. We also thank Padhraic Smyth for his guidance during the development of this paper.

REFERENCES

- Robert A Bjork. 1979. Information-processing analysis of college teaching. Educational Psychologist 14, 1 (1979), 15–23.
- [2] Nicholas J Cepeda, Harold Pashler, Edward Vul, John T Wixted, and Doug Rohrer. 2006. Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin* 132, 3 (2006), 354.
- [3] John Dunlosky, Katherine A Rawson, Elizabeth J Marsh, Mitchell J Nathan, and Daniel T Willingham. 2013. Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest* 14, 1 (2013), 4–58.
- [4] Scott Freeman, Sarah L Eddy, Miles McDonough, Michelle K Smith, Nnadozie Okoroafor, Hannah Jordt, and Mary Pat Wenderoth. 2014. Active learning increases student performance in science, engineering, and mathematics. *Proceedings of* the National Academy of Sciences 111, 23 (2014), 8410–8415.
- [5] Marissa K Hartwig and John Dunlosky. 2012. Study strategies of college students: Are self-testing and scheduling related to achievement? *Psychonomic Bulletin & Review* 19, 1 (2012), 126–134.
- [6] Jeffrey D Karpicke and Henry L Roediger. 2008. The critical importance of retrieval for learning. Science 319, 5865 (2008), 966–968.
- [7] Kayla Morehead, Matthew G Rhodes, and Sarah DeLozier. 2016. Instructor and student knowledge of study strategies. *Memory* 24, 2 (2016), 257–271.
- [8] Steve Olson and Donna Gerardi Riordan. 2012. Engage to Excel: Producing One Million Additional College Graduates with Degrees in Science, Technology, Engineering, and Mathematics. Report to the President. Executive Office of the President (2012).
- [9] Jihyun Park, Renzhe Yu, Fernando Rodriguez, Rachel Baker, Padhraic Smyth, and Mark Warschauer. 2018. Understanding student procrastination via mixture models. Proceedings of the 11th International Conference on Educational Data Mining (2018), 187–197.
- [10] Katherine A Rawson and Walter Kintsch. 2005. Rereading effects depend on time of test. Journal of Educational Psychology 97, 1 (2005), 70.
- [11] Fernando Rodriguez, Mariela J Rivas, Lani H Matsumura, Mark Warschauer, and Brian K Sato. 2018. How do students study in STEM courses? Findings from a light-touch intervention and its relevance for underrepresented students. *PLoS One* 13, 7 (2018), e0200767.
- [12] Laura J Solomon and Esther D Rothblum. 1984. Academic procrastination: Frequency and cognitive-behavioral correlates. *Journal of Counseling Psychology* 31, 4 (1984), 503.