Improving Student Performance Using Nudge Analytics

Jacqueline Feild
McGraw-Hill Education
281 Summer Street
Boston, MA USA
jacqueline.feild@mheducation.com

ABSTRACT

Providing students with continuous and personalized feedback on their performance is an important part of encouraging self regulated learning. As part of our higher education platform, we built a set of data visualizations to provide feedback to students on their assignment performance. These visualizations give students information about how they are doing compared to the rest of the class, and allow them to compare the time they spent on assignments across their courses. Included in the feedback are 'nudges' which provide guidance on how students might improve their performance by adjusting when they start or submit assignments. In order to understand what nudges to provide to students, we analyzed historical data from over 1.4 million students on over 27 million assignment submissions to find student performance trends. The data confirmed that student performance significantly decreases when assignments are started on the same day they are due and when they are submitted after the due date. We used these findings and the past and current performance of each student to display nudges relevant for them in their visualizations, highlighting actionable strategies for improving future performance.

Keywords

self-regulated learning; data visualization; data mining

1. INTRODUCTION

Self-regulation is a trait very often associated with highly effective learners [6, 1]. Feedback is an important part of the process of self-regulation, as it allows students to evaluate their performance, to decide what actions might improve their future performance and to make adjustments to their learning processes [3, 4]. Feedback can be provided in a variety of ways, but it is especially effective when it is personalized and given in near real-time. In this paper, we describe a set of data visualizations we incorporated into our higher education platform, Connect, to provide students with exactly this kind of continuous, easy to understand feedback on their assignments to encourage the development of self-regulated learning.

Specifically, these visualizations allow students to see how they are doing on assignments as soon as they are graded. In two easy-to-understand visuals they can see trends in their performance over the semester, compare their performance to the rest of the class, and compare the time they spent on each assignment across courses. In addition to this information, we use 'nudge analytics' to provide personalized messages to encourage students toward actions that might improve future performance based on patterns in historical data [2, 5]. The word 'nudge' means to encourage someone to do something, and nudge messages are an unobtrusive way to push students toward better behavior, while leaving the choice to change up to them.

To find relevant nudges, we performed exploratory data analysis on eight months of student submissions to our higher education platform, including over 1.4 million unique students and over 27 million assignments. Our goal was to find trends in the data that identify factors that lead to decreased performance for most students. In this paper, we explore the assignment submission trends by day of semester, day of week, hour of day, and started and submitted time.

2. CONNECT INSIGHT FOR STUDENTS

McGraw-Hill Education offers a teaching and learning environment, called 'Connect', for higher education. This environment allows instructors and students to manage assessments, and access ebooks and other instructional materials. As part of Connect, we built a set of visualizations called 'Insight' to help students understand their performance on assignments. These visualizations provide important feedback to students as soon as assignments are graded in a way that is easy to understand. The interactive nature allows students to make decisions about what actions might improve their future performance.

The example visualization in Figure 1 answers the question, 'How am I progressing?' and shows a student their scores on assignments in a particular class over time. The yellow trend line shows the students scores and the blue trend line shows the class average on assignments. Clicking each data point opens a right-hand panel with more details, including the nudges toward better performance if applicable. In the following sections, we will describe our analysis for determining these messages in more detail.

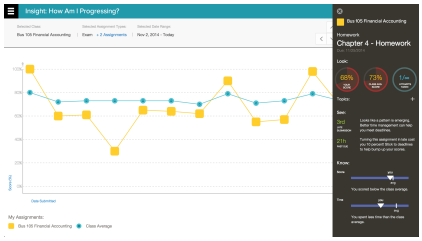


Figure 1: A visualization to answer the question 'How am I progressing?'

3. EXPLORATORY DATA ANALYSIS

In this section, we describe our analysis of eight months of historical data from assignment submissions. The goal of this analysis is to find trends in student behavior that negatively influence performance. This will help us identify the nudges that are supported by the data, and can be used to encourage students towards performance increasing behaviors.

We used historical data collected by Connect during the spring and summer semesters of 2014. This included data for 80,000 class sections taught by 29,000 instructors to 1,400,000 students. The result is over 27 million assignment submissions.

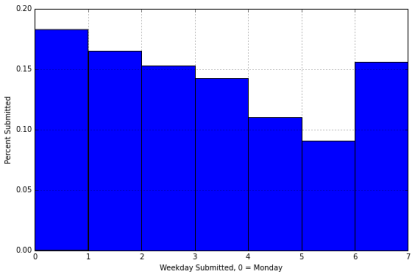
The data we used for our analysis was given to us by the Connect team from their database designed for end users, and it was not optimized for analytics. Instead, we used the existing fields for assignment submissions, including the assignment type (homework, quiz, exam, etc), start date, completion date, due date and outcome. From this data we computed a number of derived fields, including the hour of the day, day of the week and, day of the semester an assignment was submitted. We also computed the number of minutes before the due date each assignment was started and submitted.

Given these attributes, we focused our analysis on trends in assignment started and assignment submitted times. In the following sections we explore the assignment submission trends by day of semester, day of week, hour of day, and started and submitted time.

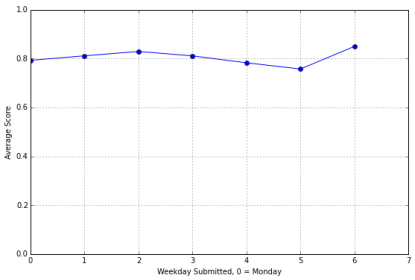
3.1 Day of the Semester

First we asked the question, does performance decrease during the semester? To start, we looked at the percent of assignments submitted on each day in our data set. This shows an interesting repeating pattern of the highest number of submissions on Monday and the lowest number of submissions on Saturday. It also shows a drop in submission volume in the middle of the spring semester, which can likely be explained by the week long spring break that occurs during this time period. Other than this decrease, submission volume remains consistent over both the spring semester and summer semester.

In order to understand student performance, we looked at the average score for assignments submitted on each day in our data set. We see a trend of decreasing performance toward the end of the spring semester (starting just before day 100). We see a similar downward trend for scores toward the end of the summer semester as well.



(a) A histogram of the percent of assignments submitted on each day of the week



(b) A plot of the average score for all submissions for each day of the week.

Figure 2

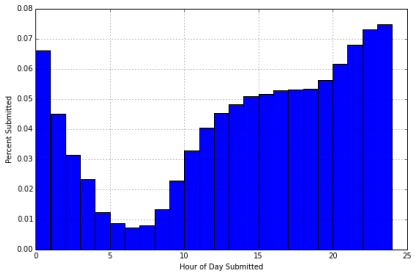
Unfortunately, there is not a clear delineation between the spring semester and the summer semester, and between the summer semester and the following fall semester, as different schools schedule classes over different time periods. Information on when classes start and end is not included in our data set, so further research is needed to confirm that this trend exists on a normalized data set.

3.2 Day of the Week

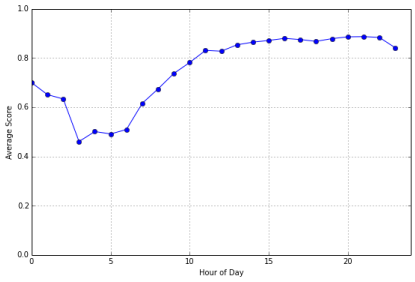
The previous analysis showed that performance decreased toward the end of the semester, but we also want to know, does performance decrease on any day of the week? Figure 2a shows the percent of assignments submitted on each day of the week. This confirms what we saw in the previous section, that the most number of assignments are submitted on Monday, while the least number of assignments are submitted on Saturday. Figure 2b shows the average score for assignments submitted on each day of the week. As expected, this shows that there is no performance advantage to submitting on a particular day of the week.

3.3 Hour of the Day

Following this analysis of scores over the semester and week, the obvious next question to explore was, does performance decrease when assignments are submitted at particular times of the day? Figure 3a shows the percent of assignment submitted during each hour of the day. This shows that most assignments are submitted between 12pm and 12am, with an increase around 8pm. While submissions do decrease in the early morning hours, there are still many submissions between 12am and 8am.



(a) A histogram of the percent of assignments submitted during each hour of the day



(b) A plot of the average score for all submissions during each hour of the day

Figure 3

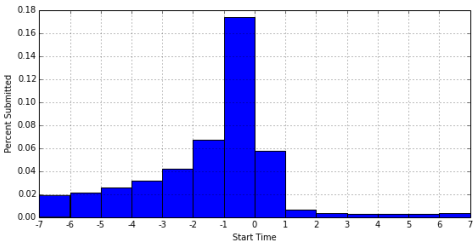
Figure 3b shows the average score for assignments submitted during each hour of the day. The average score is similar between 10am and 11pm, but steadily decreases from 11pm to 6am before increasing again. The decrease in score is significant, going from an average score of 89 at the peak hour to an average score of 46 at the lowest hour.

Unfortunately, this data represents students in many different timezones, but the the date fields are all represented in Eastern local time, where the platform servers are located. This means that we cannot draw the conclusion that submitting in the early morning hours leads to lower scores from these plots. Further work is required to obtain student time zone information and to clean the data by adjusting dates to each students local time.

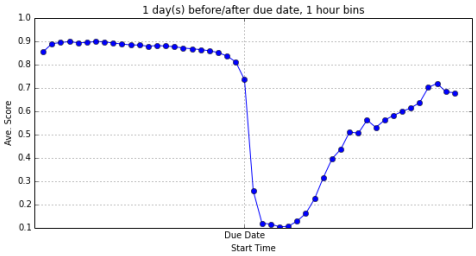
3.4 Start and Submit Time

We also looked at when a student started and submitted assignments in relation to the due date to answer the question, does assignment start time or submission time affect performance? Figure 4a is a histogram showing the percent of assignments started each day before and after the due date. The zero on the x-axis represents the deadline, so the bar between -1 and 0 represents all of the assignments that were started the same day they were due. The interesting trend in this plot is that most late assignments are started the day after the due date. This means that most late assignment could be avoided without drastic behavioral changes.

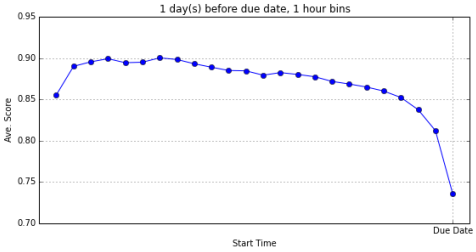
Figure 4b shows the average grade for assignments started at different points before and after the due date. The due date is in the center, and each data point to the left and right represents a 1-hour range. So the data point at the due date represents the average score of all of the assignments started



(a) A histogram of the percent of assignments started each day before and after the due date. The zero on the x-axis represents the deadline.



(b) A zoomed in version of (b) where each data point represents a 1-hour range.



(c) A zoomed in version of the left half of (c)

Figure 4

within the last hour before the due date. The point just to the left of the due date represents all of the assignments started between 1 and 2 hours ahead of the due date, and so on. In total, the plot shows one day before and after the due date. This shows that there is a decrease in average score as assignments are started closer to the deadline.

For a more detailed view, Figure 4c is a zoomed in view of Figure 4b , showing just the 24 hour window before the due date. This makes it clear that average scores significantly decrease from a high around 90 to just below 75 when started within an hour of the due date.

Plots for submit time show similar trends and are omitted due to space constraints.

The previous analysis was done using our complete data set, but we also wanted to explore whether these trends hold for each assignment type. We reproduced the plots in Figure 4 for all 14 assignment types used by our platform. We found that these trends hold for homework, quiz and exam assignments, but not all assignment types. One example where it does not hold is for LearnSmart assignments. This is most likely because Learnsmart assignments are used to drill students on a set of topics, and take a shorter period

of time to complete. Students can start them the day they are due and have plenty of time to complete satisfactorily.

4. DETERMINING DATA-DRIVEN NUDGES

We used this exploratory analysis to determine the nudge messages used in our visualizations. Based on the analysis above, conclusions could not be drawn about the day of the semester or hour of the day an assignment is submitted without further data collection and research, so these messages come from the trends seen in our exploration of start and submission time. It is clear that average scores decrease significantly as assignments are started and submitted closer to the due date and after the due date. Messages to students about when to start and when to submit are both similar in spirit, so we decided to focus our messages on starting early and avoiding submission after the due date.

We include four types of messages in our visualizations. When a student submits an assignment after the due date, they see the following message:

'Turning this assignment in late cost you <x> points! Stick to deadlines to help bump up your scores.'

and the amount of time the assignment is late is displayed in the right hand panel. When there are multiple late submissions over the semester, they will also be shown how many have been submitted late and the following additional message:

'Looks like a pattern is emerging. Better time management can help you meet deadlines.'

We also have a pair of messages focusing on starting assignments early. When students start a homework, quiz or exam within one day of the due date and they do not receive a score of 90 or better, they will receive the following message:

'Starting more than one day before the due date could result in better grades. Give yourself more time!'

If they repeatedly start assignments late, then they will see how many assignments have been started late and the following additional message:

'Late starts can lead to lower scores. Start assignments early and give yourself more time to perform better.'

These messages are designed to nudge students toward actions that will improve their performance. By providing explicit feedback about how many points they lost by submitting late, when they started assignments relative to the due date, and highlighting repeating behaviors, these message encourage students to evaluate their current actions and provide suggestions for adjusting their behavior to increase future performance on assignments.

5. CONCLUSIONS AND FUTURE WORK

In this paper we present an exploratory analysis of assignment submission data to find trends in student behavior that lead to increased performance. The data confirmed that student performance significantly decreases when assignments are started on the same day they are due and when they are submitted after the due date. We use these trends to develop data-driven nudges for students, which encourage behaviors that will help them achieve higher scores on assignments. Students see these messages when they start assignments on the same day as the due date, submit after the due date or repeatedly start or submit assignments late. These nudges are incorporated into a set of visualizations as part of our higher education platform, aimed at providing continuous, personalized feedback to students on their assignments and encouraging self-regulated learning through highlighting actionable strategies for increasing performance.

Our analysis revealed several promising avenues for future research. First, it would be interesting to understand why there are two particular assignment types that are submitted much less often than other types. This information could be used to encourage students to complete these specific assignment types or to alert instructors that these assignments are not being completed at an alarming rate and perhaps help them adjust their course to encourage completion.

We also saw potential trends in the analysis of the day of the semester assignments are completed, but we need to collect data on course start and end dates in order to clean the data set. This could lead to nudge messages reminding students to submit work as the semester progresses and scores tend to decrease. Similarly, we need to collect time zone information for each student so dates can be adjusted to local time for the analysis of the hour of the day assignments are submitted. This could lead to messages that remind students that submitting work in the early morning hours tends to lead to decreased performance.

In addition to these areas of future work, it would be interesting to do a long-term study looking at the affects of using our platform with nudge messages to understand how it affects student behaviors compared to a system that does not provide nudge messages.

6. ACKNOWLEDGMENTS

This paper is based on work supported by the McGraw-Hill Education Digital Platform Group (MHE DPG). Despite provided support, any opinions, findings, conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect positions or policies of the company.

7. REFERENCES

[1] D. L. Butler and P. H. Winne. Feedback and self-regulated learning: A theoretical synthesis. *Review of educational research*, 65(3):245–281, 1995.

[2] C. Carmean and P. Mizzi. The case for nudge analytics. *Educause Quarterly*, 33(4):n4, 2010.

[3] J. Hattie and H. Timperley. The power of feedback. *Review of educational research*, 77(1):81–112, 2007.

[4] D. J. Nicol and D. Macfarlane-Dick. Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in higher education*, 31(2):199–218, 2006.

[5] B. Wildavsky. Nudge nation: A new way to prod students into and through college. *ES Select*, 2013.

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