Relating Student Performance and Procrastination Behavior in Online Discussion Forums

Ezekiel Adriel LAGMAY[a*] & Maria Mercedes RODRIGO[a]
[a]*Ateneo Laboratory for the Learning Sciences, Ateneo de Manila University, The Philippines*
*ezekiel.lagmay@obf.ateneo.edu

Abstract: Procrastination is a behavioral feature in which a person chooses to delay a task or a decision. Academic procrastination is the tendency to postpone school-related obligations despite known negative consequences. In this paper, we examine how procrastination manifests in online discussion forum participation in a university in the Philippines during the COVID-19 pandemic. We visualize how high- and low-performing students differ in the rate at which they respond to discussion forum prompts. We also make use of association rule mining in order to determine which student behaviors are antecedents of procrastination. We find that most high-performing students tend to respond to discussion forum prompts much earlier than most low-performing students. This implies that they procrastinate less. We also found that making initial accesses or posts later or no graded posts at all makes the student at risk for poor performance.

Keywords: Procrastination, Online Learning, Philippines, COVID-19

1. Introduction

Milgram (1998 in Moonaghi & Beydokhti, 2017) defines procrastination as a behavioral feature in which a person chooses to delay a task or a decision. When applied in an educational context, academic procrastination is the tendency to postpone school-related obligations, despite known or anticipated negative consequences. There are many reasons why students may choose to procrastinate. The choice may stem from a form of sensation-seeking in which students need to work under greater time pressure. They may be distracted by media or games. They may choose to socialize instead of work. They may lack the ability to manage their time or they may feel demotivated. Regardless of the reason, studies have found that procrastination's effects are always negative.

In this paper, we discuss how procrastination manifests in online discussion forums conducted in tertiary-level classes during the COVID-19 pandemic. Specifically, we visualize how high- and low-performing students in terms of grades differ in the rate at which they respond to discussion forum prompts and we look for association rules that may point to the antecedents or precursors of procrastination. Studies of this kind may contribute to the development of criteria to detect student procrastination and can help in the design of educational interventions to mitigate this behavior. The research questions for this study are as follows:

1. How long does it take for high- and low-performing students to make an initial access to a course discussion forum, whether graded or ungraded?
2. How long does it take for high- and low-performing students to make an initial post to a course discussion forum, whether graded or ungraded?
3. What is the relationship between the average days elapsed to initial access or post and the performance (estimated overall average grade) of a student?

2. Review of Related Literature

Prior literature has found associations between procrastination and poor learning outcomes in general. Studies that investigate student participation and performance in online classes provide corroborating evidence for these conclusions. In what claims to be the first empirical study for the relationship between procrastination and participation in online discussion forums, Michinov and colleagues (2011) found that high procrastinators were less successful than low procrastinators partly due to the former's lack of participation in online discussion forums. Jones and Blankenship (2021) investigated the relationship between submission time and academic grades before, on, and after a submission. They found that earlier submission dates tend to be positively related with higher grades.

Studies on student participation during the COVID-19 pandemic continued to provide evidence that procrastination negatively correlates with learning. They also provided evidence of the relationship between procrastination and other characteristics and traits that in turn have implications on how well students cope with their academics. Procrastination was less common, for example, among students who were intrinsically motivated and who were capable of self-regulated learning strategies such as time management and goal-setting (Pelikan, Lüftenegger, Holzer, Korlat, Spiel, & Schober, 2021). Procrastination was more common among students who perceived themselves as having lower competence. Students with high levels of social support from roommates, partners, and family tended to report lower levels of academic procrastination (Liimatta, 2021). Students who were present-biased, i.e. had a preference for a small reward now rather than a large reward later, procrastinated possibly because studying with peers kept them focused and committed (De Paola, Gioia, & Scoppa, 2022). In the absence of such social pressures, they were more easily distracted by social media, games, and other applications.

Researchers use a number of learning analytics methods to arrive at these relationships. In the study conducted by Goda et al. (2014), the aim was to extract learning behavioral types in e-learning and discover the relationships between the different learning types and learning outcomes. After determining the seven learning behavioral types (including procrastination) based on the weekly completion rates of the learning materials and weekly-accumulated completion rates in the first phase, the researchers proceeded with correlating these learner behavior types with learning performance indicators (in this case, the scores of TOEIC-IP) using ANOVA (Goda, Yamada, Kato, Matsuda, Saito, & Miyagawa, 2014). The results showed that students who fall into the learning habit or chevron types significantly scored higher on the TOEIC-IP than the procrastination type (Goda et al., 2014).

In another study conducted by Cerezo et al. (2017), the proponents used Moodle logs to extract variables related to procrastination including those related to the number of days the students wait to check each assignment, task, forum subject, and theoretical content, and those related to the number of days that they took to hand in their task and post their opinion (Cerezo, Esteban, Sánchez-Santillán, & Núñez, 2017). These characteristics for procrastination were chosen as the proponents wanted to observe "the students' behavioral patterns before the homework deadline and not solely considering late or absent submissions" (Cerezo et al., 2017). Class Association Rules (Predictive Apriori) were then applied to the data after it was discretized using equal-width for the antecedent variables and a custom method based on the Spanish grading system for the performance (class/consequent variable) (Cerezo et al., 2017). The resulting association rules showed that, in general, "evidence of procrastination in the antecedents leads to poor performance, and signs of successful time management end up with satisfactory achievement" (Cerezo et al., 2017).

This study follows the methods used by Cerezo and colleagues (2017) in order to determine how high- and low-performers from a university in the Philippines differ in their procrastination behavior. Furthermore, we use association rule mining to find the antecedents of procrastination, increasing the likelihood of early detection.

3. Methodology

3.1 Description of Dataset

The dataset consisted of Canvas log files collected from a university in Quezon City, Metro Manila, Philippines during the first semester of Academic Year (AY) 2021-2022 (August 26, 2021 to December

18, 2021) (De Leon, 2021). The process of selecting the logs to analyze was as follows:

- The logs should come from a course discussion forum topic which has a definite classification of whether it is graded or not.
- The logs, whether posts or accesses, are generated by students whose overall grades can be estimated from 6 or more enrolled courses (ACU International, 2021; Ateneo de Manila University, 2021).
- The logs should come from a course wherein the list of students could be determined from the asset accessed logs, since data for the list of students enrolled in a course is not readily available from the Canvas log files.
- The logs occurred between August 26, 2021 to December 18, 2021 (in UTC).

3.2 Data Preprocessing

The course discussion forum log files are first grouped according to whether it is an Access or a Post. Accesses are considered in addition to Posts in order to account for the possibility of "lurkers" (i.e. those who just access the course discussion forums but not to post) (Guzdial & Carroll, 2002; Nandi, Hamilton, Harland, & Warburton, 2011; Salmon, 2003). For each of the categories, the data is split based on whether the course discussion forum topic each log pertains to is graded or ungraded. There was a total of 3,956,866 course discussion forum access logs and 206,767 course discussion forum post logs analyzed, of which 1,311,437 logs and 83,923 logs are accesses and posts, respectively, to graded course discussion forums. Hence, there are a total of four analyses that were made: Accesses-Ungraded, Accesses-Graded, Posts-Ungraded, and Posts-Graded.

For each of the analyses, the data preprocessing procedure starts by aggregating and getting the earliest date of post/access per course ID, course discussion topic ID, and user ID, which is then subtracted to the overall earliest date of post/access per course ID and course discussion topic ID. In order to accommodate for the possibility that the student may not have posted at all in a particular course discussion forum topic, an imputation of values was done wherein per course, a cross-product of the complete list of topics and students was performed and then merged with the observed data to generate a complete dataset, with the students who did not post on a particular course discussion topic given a value of 115 for the days elapsed from the first post/access for that course discussion topic. 115 is the total number of days comprising the 1st Semester AY 2021-2022 period, and since the first day, August 26, 2021, is marked as Day 0, Day 115 pertains to the date of December 19, 2021, the day after the last day of 1st Semester AY 2021-2022 (December 18, 2021).

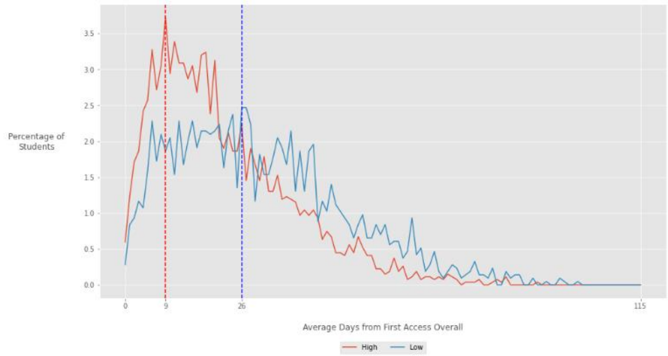
3.3 Statistical Analysis

For each of the four analyses, two sub-analyses are made. The "Overall" sub-analysis involves taking the average of the days elapsed from the date of the first post/access for each user regardless of the course, and then getting the performance ranking of each user (High if estimated overall average grade is greater than or equal to 87% or Low otherwise) ("Academic Grading in the Philippines", 2022). A total of 2,688 high-performing students and 2,148 low-performing students were included in the Accesses-Ungraded and Posts-Ungraded analyses, of which 2,657 high-performing students and 2,116 low-performing students were part of the Accesses-Graded analysis, and of which 2,655 high-performing students and 2,113 low-performing students are included in the Posts-Graded analysis. The percentage of the total number of students who made a post/access per ranking is then obtained per days elapsed since the initial post/access. On the other hand, the "Per Course" sub-analysis is similar except that the average of the days elapsed from the date of the first post/access for each user is first calculated per course, and then the values per course are then averaged together per user. For both of the two sub-analyses, another dataset is generated by bucketing the averaged days elapsed from initial post/access per user. These resulting datasets (8 in total, each pertaining to a particular Posts/Accesses-Graded/Ungraded-Overall/Per Course combination, with a total of 4,768 students included) are then used for the association rules analysis procedure using the Predictive Apriori Algorithm in Weka 3.8.6, with the overall performance of the student as the consequent variable (Cerezo et al., 2017). Two association rules analysis results were generated, one for the "Overall" sub-analysis and another for the "Per Course" sub-analysis.

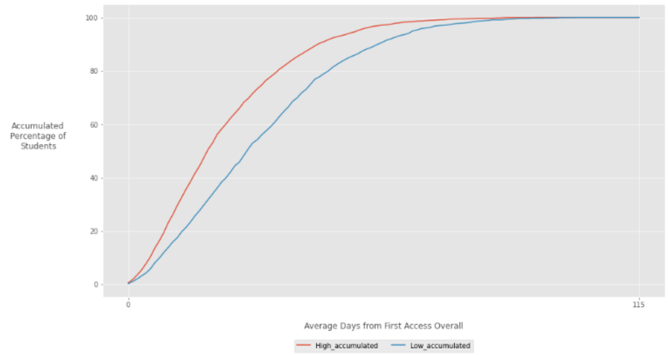
4. Results and Discussion

4.1 Accesses-Ungraded

In both Figures 1a and 2a, it could be clearly seen that the majority of the high-performing students begin accessing the ungraded course discussion forums earlier than their low-performing counterparts. In Figure 1a, approximately 3.72% of the high-performing students begin accessing the ungraded course discussion forums 9 days from the initial access overall while around 2.47% of the low-performing students begin accessing 26 days from the initial access overall. In Figure 2a, on a per course perspective, approximately 3.57% of the high-performing students begin accessing the course discussion forums 14 days from the initial access, while around 2.61% of the low-performing students begin accessing the course discussion forum 20 days from the initial access. Also, as the days elapsed from the initial access increases, the percentage of low-performing students begin to outnumber the percentage of high-performing students. In the accumulated percentage graphs shown in Figures 1b and 2b, the high-performing students reach the peak first before the low-performing students.

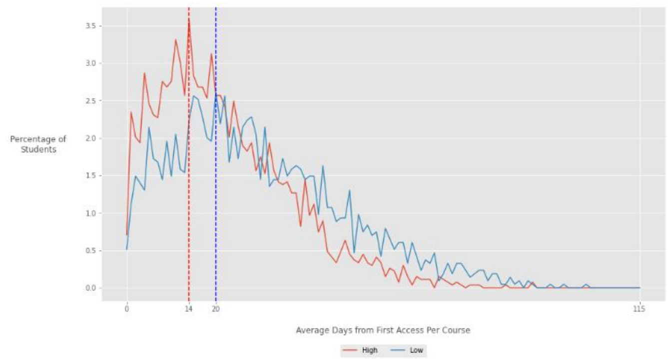


a. Raw Percentage

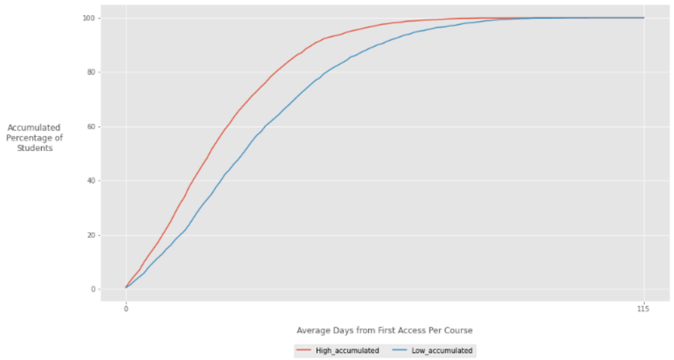


b. Accumulated Percentage

Figure 1. The Percentage of Students per Ranking Visualized According to the Average Days Elapsed From First Access Overall – Ungraded.



a. Raw Percentage

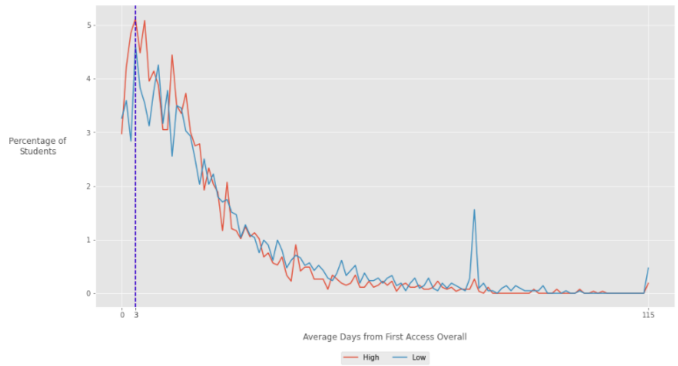


b. Accumulated Percentage

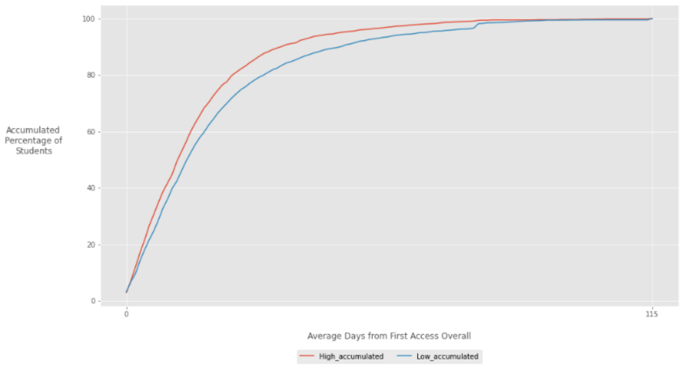
Figure 2. The Percentage of Students per Ranking Visualized According to the Average Days Elapsed From First Access Per Course – Ungraded.

4.2 Accesses-Graded

In both Figures 3a and 4a, it could be clearly seen that the majority of both high-performing and low-performing students begin accessing the graded course discussion forums on the same day – Day 3. In Figure 3a, approximately 5.12% of the high-performing students and 4.63% of the low-performing students begin accessing the graded course discussion forums overall on Day 3. In Figure 4a, on a per course perspective, approximately 5.72% of the high-performing students and 5.25% of the low-performing students begin accessing the course discussion forums 3 days from the initial access. Also, both Figures 3a and 4a do not suggest a significant difference between the high-performing and the low-performing students regardless of the number of days elapsed from the initial access. But when considering the accumulated percentage graphs on Figures 3b and 4b, the high-performing group reaches close to the peak first before the low-performing group.

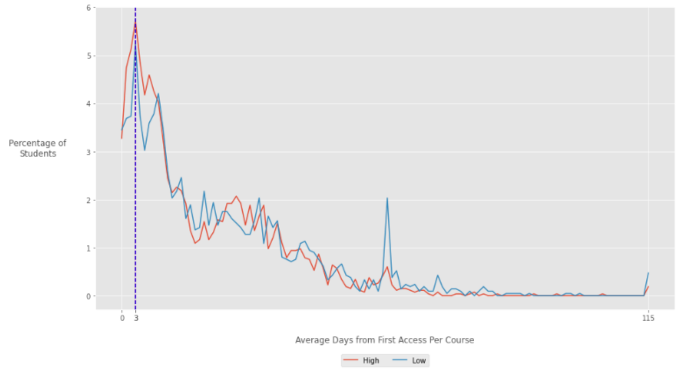


a. Raw Percentage

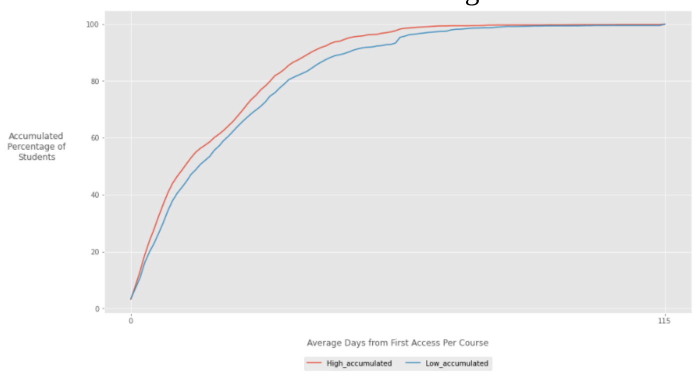


b. Accumulated Percentage

Figure 3. The Percentage of Students per Ranking Visualized According to the Average Days Elapsed From First Access Overall – Graded.



a. Raw Percentage

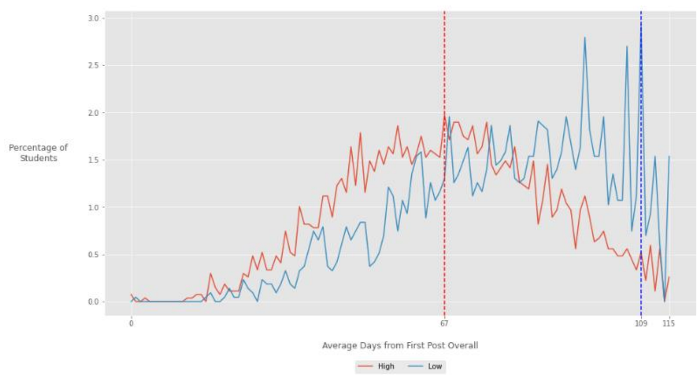


b. Accumulated Percentage

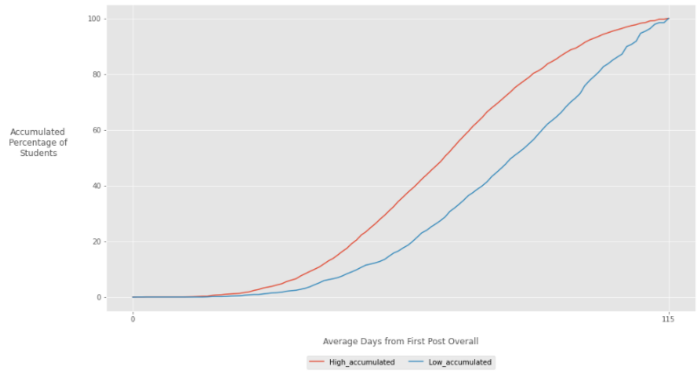
Figure 4. The Percentage of Students per Ranking Visualized According to the Average Days Elapsed From First Access Per Course – Graded.

4.3 Posts-Ungraded

In both Figures 5a and 6a, it could be clearly seen that the majority of the high-performing students begin posting to the ungraded course discussion forums earlier than their low-performing counterparts. In Figure 5a, approximately 1.97% of the high-performing students begin posting to the ungraded course discussion forums 67 days from the initial post overall while around 2.93% of the low-performing students begin posting 109 days from the initial post overall. In Figure 6a, on a per course perspective, approximately 2.72% of the high-performing students begin posting to the course discussion forums 60 days from the initial post, while around 2.09% of the low-performing students begin posting to the course discussion forum 88 days from the initial post. Also, as the days elapsed from the initial post increases, the percentage of low-performing students begin to outnumber the percentage of the high-performing students. When considering the accumulated percentage graphs on Figures 5b and 6b, the high-performing group reaches close to the peak first before the low-performing group.

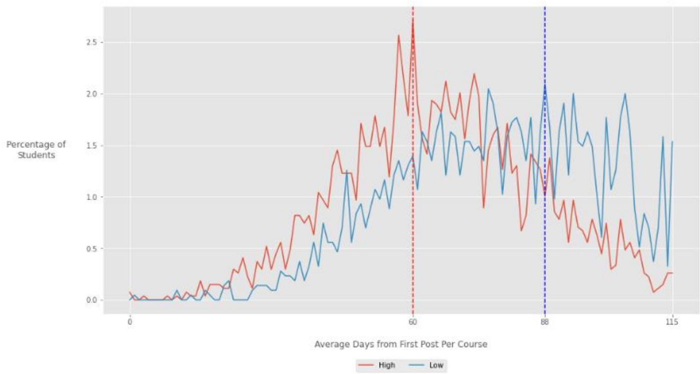


a. Raw Percentage

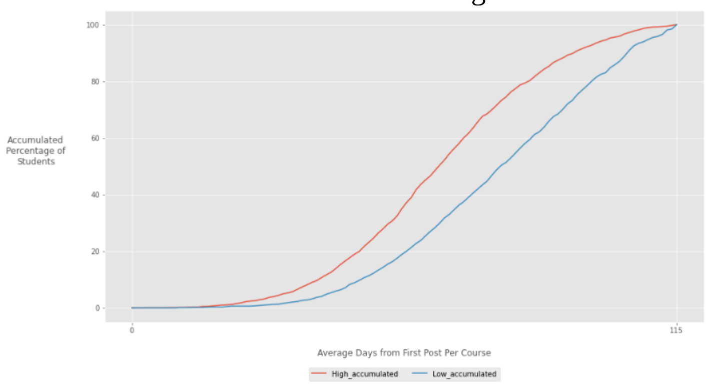


b. Accumulated Percentage

Figure 5. The Percentage of Students per Ranking Visualized According to the Average Days Elapsed From First Post Overall – Ungraded.



a. Raw Percentage

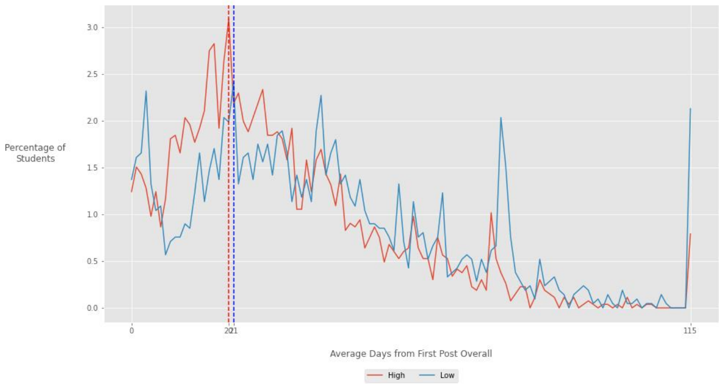


b. Accumulated Percentage

Figure 6. The Percentage of Students per Ranking Visualized According to the Average Days Elapsed From First Post Per Course – Ungraded.

4.4 Posts-Graded

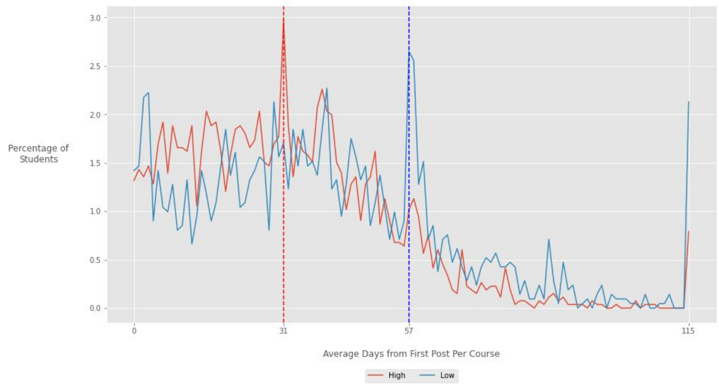
In both Figures 7a and 8a, it could be clearly seen that the majority of the high-performing students begin posting to the graded course discussion forums earlier than their low-performing counterparts. In Figure 7a, approximately 3.09% of the high-performing students begin posting to the graded course discussion forums 20 days from the initial post overall while around 2.41% of the low-performing students begin posting 21 days from the initial post overall. In Figure 8a, on a per course perspective, approximately 2.98% of the high-performing students begin posting to the course discussion forums 31 days from the initial post, while around 2.65% of the low-performing students begin posting to the course discussion forum 57 days from the initial post. Also, both Figures 7a and 8a do not suggest a significant difference between the high-performing and the low-performing students regardless of the number of days elapsed from the initial access. What is more alarming, however, is the relatively high percentage of low-performing students who did not post at all to a graded discussion forum (Day 115). When considering the accumulated percentage graphs on Figures 7b and 8b, the high-performing group once again reaches close to the peak first before the low-performing group.



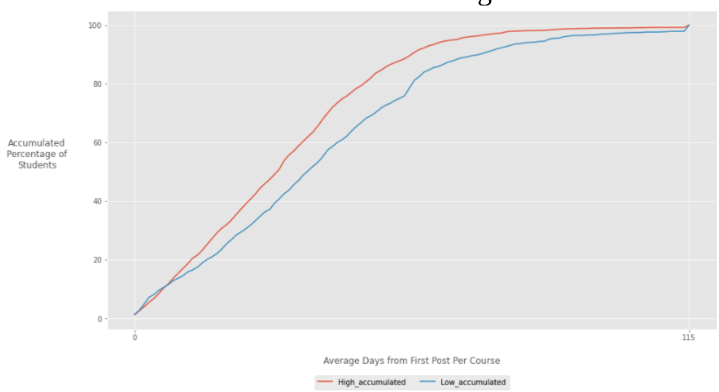
a. Raw Percentage

b. Accumulated Percentage

Figure 7. The Percentage of Students per Ranking Visualized According to the Average Days Elapsed From First Access Overall – Graded.



a. Raw Percentage



b. Accumulated Percentage

Figure 8. The Percentage of Students per Ranking Visualized According to the Average Days Elapsed From First Access Per Course – Graded.

4.5 Association Rules Analysis

In Table 1, when looking at an overall level, the high-performing students post/access the online course discussion forums earlier than the low-performing students. Furthermore, those who fail to post in a graded online course discussion forum are more likely to perform poorly. On the other hand, when looking on a per course level (Table 2), the same pattern also emerges in that low-performing students make their initial post/access much later compared to the high-performing students, and that those who did not make a single post at all in a graded course discussion forum are all the more likely to have low performance.

Table 1. *Top 5 Association Rules Analysis Results for the Averaged Days Elapsed from First Access or Post Overall*

Access Ungraded	Access Graded	Posts Ungraded	Posts Graded	Performance	Accuracy
N/A	N/A	Day 25 to 30	Day 7 to 12	High	0.99
N/A	Day 73 to 78	N/A	None	Low	0.99
Day 1 to 6	N/A	Day 13 to 18	N/A	High	0.98
Day 43 to 48	N/A	Day 61 to 66	N/A	High	0.98
N/A	N/A	Day 31 to 36	Day 19 to 24	High	0.97

Table 2. *Top 5 Association Rules Analysis Results for the Averaged Days Elapsed from First Access or Post Per Course*

Access Ungraded	Access Graded	Posts Ungraded	Posts Graded	Performance	Accuracy
N/A	N/A	Day 109 to 114	None	Low	0.99
N/A	N/A	Day 103 to 108	None	Low	0.99
Day 1 to 6	N/A	Day 19 to 24	N/A	High	0.98
Day 79 to 84	N/A	N/A	N/A	Low	0.98
Day 67 to 72	N/A	N/A	Day 55 to 60	Low	0.96

5. Conclusion, Limitations, and Further Studies

Overall, the visualizations showed that, regardless whether the course discussion forum is graded or not, a high percentage of the high-performing students made posts and accesses earlier than the low-performing students. When the percentages are accumulated, almost the entire high-performing group made their initial access/posts before the entire low-performing group does. On the other hand, the association rules showed that performing initial accesses or posts later, or even not post in a graded online course discussion forum at all, is an indicator of possible poor performance. These features may be helpful in identifying students who are likely to procrastinate and may cue intervention.

This study, though, is subject to several limitations. Student performance is defined as the estimated overall average grade the student has received for all courses, and course contents were not taken into consideration. There are other indicators of student performance or behavioral factors such as fear, perfectionism, and motivation. Future studies can take these into consideration.

Detecting procrastination behavior is the first step in mitigating it. Suggested interventions from prior research include but are not limited to making forum participation compulsory (Gafni & Geri, 2010), converting large, one-off course requirements into smaller requirements spread throughout the academic term (Kang & Zhang, 2020), providing scaffolding to help extreme procrastinators learn more adaptive learning strategies (Cerezo, Sánchez-Santillán, Paule-Ruiz, & Núñez, 2016), and greater social support (Liimatta, 2021). Future research could try to implement these strategies and measure their effects on student online participation.

Acknowledgements

The authors would like to thank the Ateneo Laboratory for the Learning Sciences, Ateneo Research Institute for Science and Engineering (ARISE), and Accenture for the funding and support needed for this research. The authors would also like to thank Deni Jaramillo and Miguel Saavedra for their assistance in setting up the necessary servers for the collection of Canvas data.

References

Academic Grading in the Philippines. (2022). Retrieved from: https://en.wikipedia.org/wiki/Academic_grading_in_the_Philippines

ACU International. (2021). Ateneo de Manila University The Philippines. Retrieved from: https://www.studentportal.acu.edu.au/-/media/acu/portal/pdfs/acu-international/fact-sheets/asia/philippines---ateneo-de-manilla-university.pdf?la=en&hash=BF2E8C89D34D296E2DF31248118E166E&hash=BF2E8C89D34D296E2DF31248118E166E

Ateneo de Manila University. (2021). The Loyola Schools Undergraduate Student Handbook 2021. Retrieved from: https://bit.ly/LSStudentHandbook2021

Cerezo, R., Sánchez-Santillán, M., Paule-Ruiz, M. P., & Núñez, J. C. (2016). Students' LMS Interaction Patterns and Their Relationship with Achievement: A Case Study in Higher Education. *Computers & Education*, *96*, 42-54.

Cerezo, R., Esteban, M., Sánchez-Santillán, M., & Núñez, J. C. (2017). Procrastinating Behavior in Computer-Based Learning Environments to Predict Performance: A Case Study in Moodle. *Frontiers in Psychology: Educational Psychology*, *8*(1403).

De Leon, M. M. (2021). Ateneo de Manila University Loyola Schools Office of the Registrar – Academic Calendar for the First and Second Semester, School Year 2021-2022. Retrieved from: https://sites.google.com/ateneo.edu/ls-one/ls-memos/memo-archives#h.augtcnqt5au4

De Paola, M., Gioia, F., & Scoppa, V. (2022). Online Teaching, Procrastination and Students' Achievement: Evidence from COVID-19 Induced Remote Learning.

Gafni, R., & Geri, N. (2010). The Value of Collaborative E-Learning: Compulsory Versus Optional Online Forum Assignments. *Interdisciplinary Journal of E-Learning and Learning Objects*, *6*(1), 335-343.

Goda, Y., Yamada, M., Kato, H., Matsuda, T., Saito, Y., & Miyagawa, H. (2014). Procrastination and Other Learning Behavioral Types in E-Learning and Their Relationship with Learning Outcomes. *Learning and Individual Differences*, *37*(2015), 72-80.

Guzdial, M., & Carroll, K. (2002). Explaining the Lack of Dialogue in Computer-Supported Collaborative Learning. In *The Computer Supported Collaborative Learning Conference 2002*.

Jones, I. S., & Blankenship, D. C. (2021). Year Two: Effect of Procrastination on Academic Performance of Undergraduate Online Students. *Research in Higher Education Journal*, *39*.

Kang, X., & Zhang, W. (2020). An Experimental Case Study on Forum-Based Online Teaching to Improve Student's Engagement and Motivation in Higher Education. *Interactive Learning Environments*, 1-12.

Liimatta, P. O. (2021). *The Relationship between Living Situation, In Person and Online Social Support and Academic Procrastination During the COVID-19 Pandemic* (Bachelor's thesis, University of Twente).

Michinov, N., Brunot, S., Le Bohec, O., Juhel, J., & Delaval, M. (2011). Procrastination, Participation, and Performance in Online Learning Environments. *Computers & Education*, *56*(1), 243-252.

Moonaghi, H. K., & Beydokhti, T. B. (2017). Academic Procrastination and Its Characteristics: A Narrative Review. *Future of Medical Education Journal*, *7*(2), 43-50.

Nandi, D., Hamilton, M., Harland, J., & Warburton, G. (2011). How Active are Students in Online Discussion Forums? In *13th Australasian Computing Education Conference Proceedings* (pp. 125-134). Perth, Australia: Australian Computer Society, Inc.

Pelikan, E. R., Lüftenegger, M., Holzer, J., Korlat, S., Spiel, C., & Schober, B. (2021). Learning During COVID-19: The Role of Self-Regulated Learning, Motivation, and Procrastination for Perceived Competence. *Zeitschrift für Erziehungswissenschaft*, *24*(2), 393-418.

Salmon, G. (2003). *E-tivities: The Key to Active Online Learning*. London: Kogan Page.

Yang, Y., Hooshyar, D., Pedaste, M., Wang, M., Huang, Y.-M., & Lim, H. (2020). Predicting Course Achievement of University Students Based on Their Procrastination Behaviour on Moodle. *Soft Computing*.