Investigating Student Participation and Performance in Calculus I Courses that Utilize Standards-Based Grading

Abstract

In this study, we analyzed student participation and success in a college-level calculus I course that utilized standards-based grading. By defining student "buy-in" as the level to which students participate in this class structure, we were able to use a clustering algorithm that revealed multiple groupings of students that were distinct based on activity throughout the semester. Additionally, we analyzed student progress, defined as the number of graded activities successfully completed each week. We found that students who progressed steadily throughout the semester, and thus had lower variability in the number of completed activities per week, tended to receive a higher overall grade. Students whose progress was less consistent, and thus exhibited higher variability in weekly activities completed, tended to receive a lower grade. Overall, this shows implications for the pay-off of buying into the method of standards-based grading and succeeding in a course.

1 Introduction

Standards-based grading (SBG) is a nontraditional grading method that is beginning to become more popularized in the United States. Instead of receiving grades on exams and homework based on points and partial credit, students are presented with a list of standards. These standards are directly linked to student learning outcomes, and a student's grade is determined by the number of standards met. The standards are designed to demonstrate understanding of specific topics and, typically, each standard builds on the concepts from previous standards. The standards are graded pass/fail without the possibility of partial credit. To counteract the loss of partial credit, the students are given multiple attempts at each standard in hopes that they will learn from previous mistakes. This type of grading is typically implemented at an elementary or middle school level and is just starting to move into upper-level education such as high schools and colleges.

Standards-based grading requires that students adjust their learning approach and that professors and teachers change their teaching methods. The shift in learning can make it difficult for some students to "buy-in" and consistently work hard to achieve a good grade. According to research done by Lewis (2019), standards-based grading has been found to affect students' stress levels, when it comes to mathematics classes, in different ways. Students typically reported that they felt less anxious when it came to exams and homework assignments because they knew they could retry them again; however, student stress seemed to increase over time as the number of standards students were required to complete began to compile into a large workload.

In education, there has been a disparity between the experiences of high-achieving and low-achieving students (defined by traditional methods of assessment). SBG is an educational system that has the potential to motivate a broader range of students than the traditional system. Meyer et al. (2009) conducted a study in New Zealand with participants of senior secondary students from 20 nationally representative secondary schools. The study evaluates correlations between self-reported motivation and achievement on the NCEA (an SBG system for secondary students). The findings indicate significance between the designs of a standards-based assessment system and student attitudes and achievement. This study also highlights the need for long-term studies investigating student motivation and achievement.

Our research attempts to determine if college students, particularly in calculus I classes, participate and achieve better grades depending on how much they buy-in to the standards-based grading method. Our definition of buy-in depends on a particular student's ability to pass and willingness to retry failed standards throughout the semester. Furthermore, buy-in is quantified by looking at the amount of variability in each student's progress throughout the semester. We started with a simple analysis of our data and used our definition of buy-in to identify how a bought-in student would look different from a student who was not bought in to the SBG method. We analyzed our data more extensively by using K-means clustering to break the students into groups based on their level progression throughout the semester. This clustering was used to identify patterns of behavior among students. Next, logistic regression was utilized to investigate how variance in level progression relates to a student's overall grade.

2 Data

2.1 Data Collection

Our data were collected from three different sections of calculus I classes in the spring and fall semesters of 2021 taught by PROFESSOR at COLLEGE. Each student's progress was recorded throughout the semester and included their checkpoint attempts, class activities, and final grades. For each standard given, students were allotted a set amount of time, typically a week, to finish the standard completely and turn it in. They would then receive either a pass or fail on the standard. If a student failed a standard, they were then given another week to reattempt it. These reattempts were given until a student passed the checkpoint, missed the deadline, or the semester ended. Class activities included group work assignments and smaller tasks that would be graded based on completion. Student attendance was also taken into account for grading. At the beginning of the semester, students were given a survey asking about how many years it had been since they had taken a math course and their expectations for the course. All of these factors were taken into account for our initial data analysis. The level numbers corresponding to activities completed is shown in Table 1.

Level	CORE Checkpoints	Total Checkpoints	Preview Activities	Team Activities	Webwork Assignments
25	8	23	22	45	12
24	8	22	21	43	12
23	7	21	20	41	11
22	7	20	19	40	10
21	7	19	18	38	10
20	6	18	18	36	9
19	6	17	17	34	9
18	6	17	16	32	8
17	5	16	15	31	8
16	5	15	14	29	7
15	5	14	13	27	7
14	4	13	12	25	6
13	4	12	11	23	6
12	4	11	10	21	5
11	4	10	9	19	5
10	3	9	9	17	4
9	3	8	8	15	4
8	2	7	7	13	3
7	2	6	6	11	3
6	1	5	5	9	2
5	1	4	4	7	2
4	0	3	4	5	1
3	0	2	3	3	0
2	0	1	1	1	0
1	Started the Course!				

Table 1: Activities completed to level conversion.

Students' final grades were calculated based on the level they had achieved by the end of the course. These levels were determined by the number of checkpoints passed, class activities, and participation in various other activities. A student may fail to advance in level by a number of different factors such as missing a checkpoint or not attending class. Table 2 illustrates how a student's final level achieved was converted to their final grade for the course.

Grade	Minimum Level	Grade	Minimum Level	Grade	Minimum Level
A	25	B-	21	D+	17
A-	24	C+	20	D	16
B+	23	C	19	D-	15
B	22	C-	18	F	14

Table 2: Level to grade conversion.

2.2 Defining Buy-in

To begin data analysis, we assigned each student a six-number ID to protect their identities. We then took the numerical data that were collected as well as the students' responses to the pre-semester survey and compiled them into a summary dataset for analysis.

One of our main concerns was the definition of buy-in for a student. We decided to define a bought-in student as having consistent progress throughout the semester without much variability. A student who was not bought in would have periods of time where their progress stopped completely or was not moving at a consistent rate. The comparison of two example students, one being bought-in and the other not, is demonstrated in Figure 1.

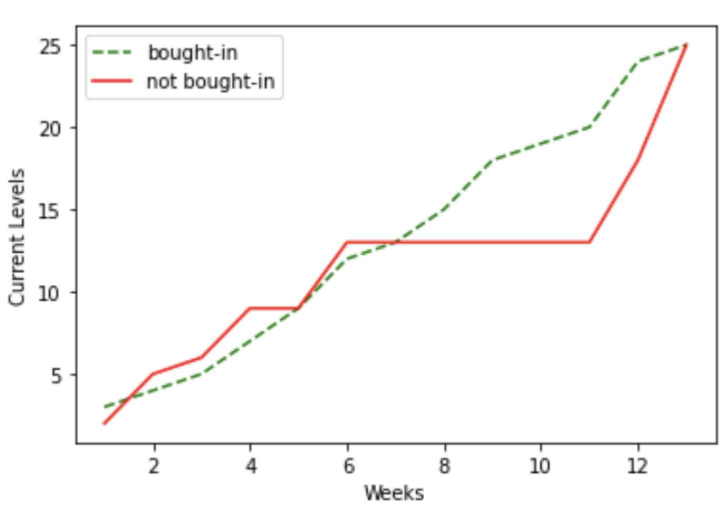


Figure 1: Example of a bought-in vs a non bought-in student.

We wanted to further investigate the parameters and qualities of bought-in students compared to students who were not bought in. To analyze how the average rate of progress changes throughout the semester, we graphed the average levels progressed each week for students in the Spring semester. This is shown in Figure 2.

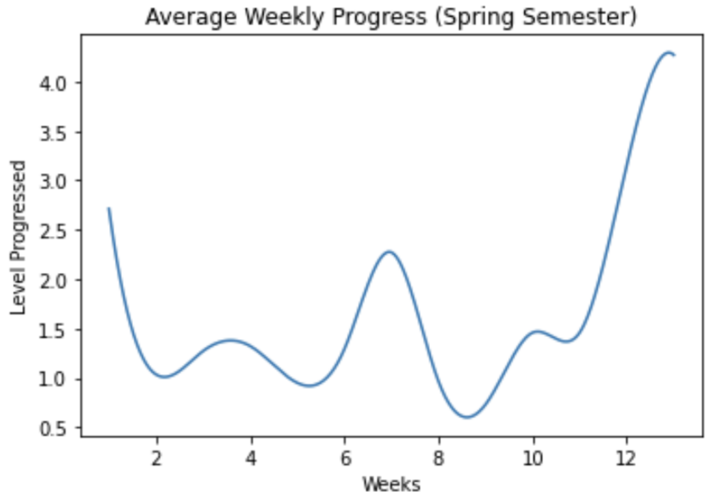


Figure 2: Distribution of average weekly level progress.

As it can be seen, there were large jumps in level progress during the first week of classes, the week after a break, and during the last week of classes. This jump was especially prominent during the last week when students were beginning to scramble to raise their grade in the class, which resulted in an average of four levels being progressed in a single week. Since the students were typically having spikes in progress at distinct times, we decided to utilize k-means clustering and logistic regression to pinpoint these specific patterns and identify how they could be linked to students' final grades in the course.

3 Identifying Different Patterns in Students' Progress

We wanted to identify different patterns of behaviors between students in terms of weekly progress. In this section, we present a way to construct time series from students' activities recorded throughout the semester and use a clustering algorithm to group the students based on the shape of their corresponding time series.

3.1 Representing Students' Progress as Time Series

A time series is a one-dimensional vector indexed by time interval. In order to view the students' progress throughout the semester, we calculated each student's weekly levels and used the data to construct a corresponding time series. Each student is represented by only one time series, and each entry in the time series corresponds to the student's level at that given week.

Week	Level	Week	Level	Week	Level
1	3	5	13	9	17
2	5	6	15	10	19
3	7	7	16	11	20
4	9	8	16	12	22

Table 3: Example of a student's weekly progress.

Table 3 shows the calculated weekly progress of a hypothetical student, student A. From their performance record, we can construct their progress time series as:

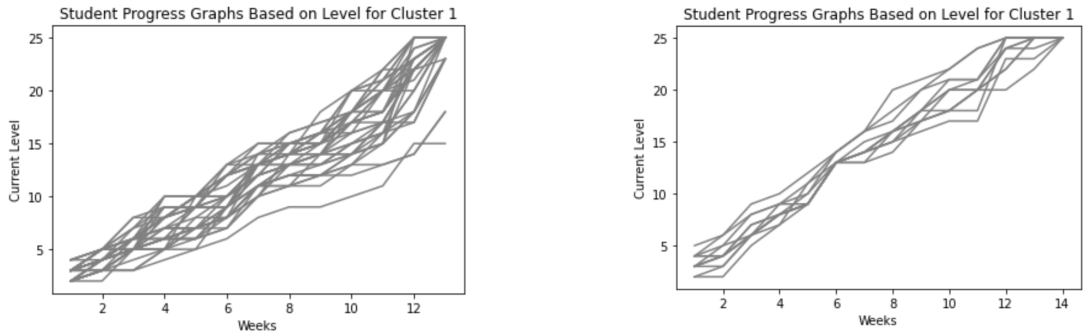
$$\vec{t}_A = (3, 5, 7, 9, 13, 15, 16, 16, 17, 19, 20, 22)$$

Replicating the same process for all students in our dataset, we obtained the final result of 88 time series in total. Among those, there are 59 time series from the spring semester and 29 time series from the fall semester. Due to the COVID-19 pandemic, the spring semester was one week shorter than the fall semester. As a result, the spring semester's time series contains 12 entries, while the fall semester's contains 13 entries.

3.2 Clustering Times Series

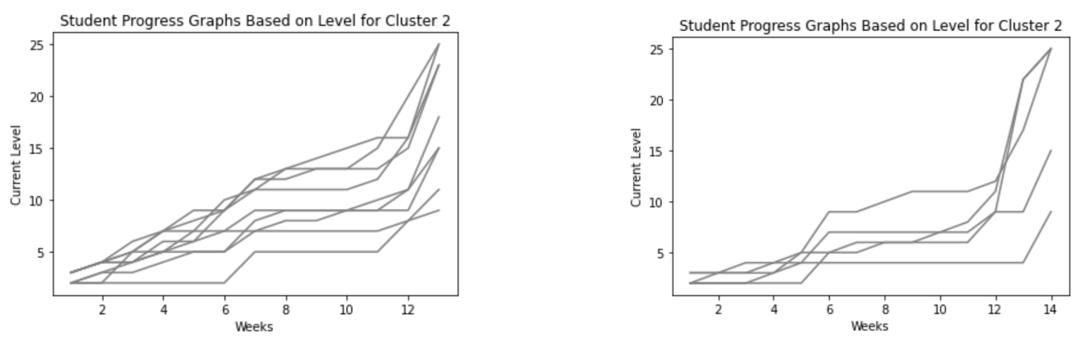
Next, we employed k-mean clustering, an unsupervised clustering algorithm, to group students into different clusters based on the shape of their corresponding time series. The metrics we used to perform clustering is dynamic time warping. First, we normalized the time series using mean-standard scalar to transform all series to have a mean equal to 0 and a standard deviation equal to 1. This ensured that the time series were clustered based on their shape.

Silhouette score/coefficient is a metric used to assess the performance of the clustering algorithm. We calculated silhouette scores for different numbers of clusters (n) to choose the best n for our problem. The value of silhouette scores can range from -1 to 1, with 1 denoting that all data points within the cluster are compact and far away from other clusters, and -1 denoting the opposite. The value of 0 shows overlapping clusters. After performing silhouette analysis on various numbers of clusters (Table 4), we decided to choose the number of clusters equal to three. The results of the clustering algorithm is presented below:



(a) Spring Semester (b) Fall Semester

Figure 3: Cluster 1 (spring and fall semester).



(a) Spring Semester (b) Fall Semester

Figure 4: Cluster 2 (spring and fall semester).

Number of Cluster	Silhouette Score
2	0.31
3	0.37
4	0.12
5	-0.032

Table 4: Results of silhouette analysis.

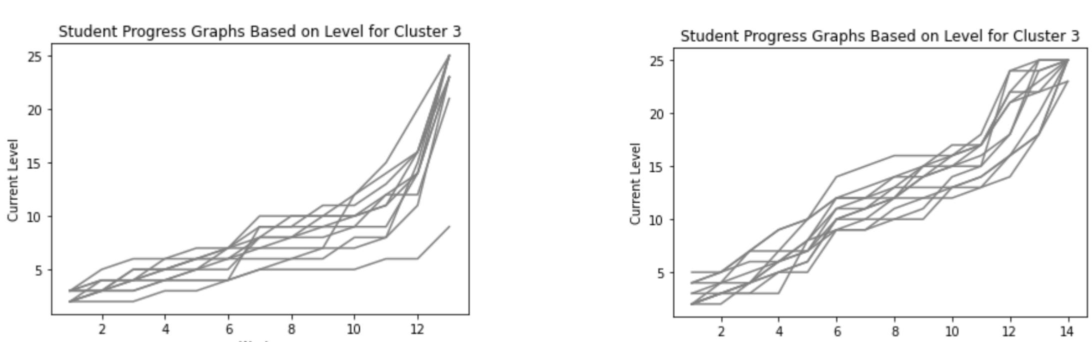
Figure 3 shows the time series clustered into group 1. As it can be observed from the graphs, this group's main characteristic is early and consistent progress throughout the semester. This group also contains the highest number of students overall, making up 58% of the entire dataset. We call this group "The Consistents".

Figure 4 shows the time series clustered into group 2. As it can be observed from the graphs, this group's main characteristic is a lack of progress during the middle weeks of the semester. Some students exhibited no or little progress for up to eight weeks. This group also contains the minority of students overall, making up 15.8% of the entire dataset. We refer to this group as "The Sluggers".

Figure 5 shows the time series clustered into group 3. As it can be observed from the graph, this group's main characteristic is slow progress for most parts of the semester, followed by a sudden burst of activities during the last two weeks of the semester. This group contains 26.8% of the entire dataset. We call this group "The Procrastinators".

3.3 Patterns of Progress and Final Grades

As shown in the previous subsection, there exists distinct patterns of progress among the students. In this subsection, we want to examine whether students with different patterns of progress are likely to receive



(a) Spring Semester (b) Fall Semester

Figure 5: Cluster 3 (spring and fall semester).

different final grades.

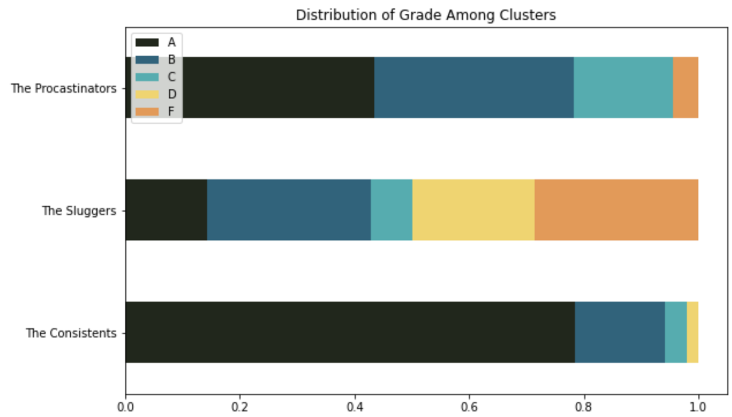


Figure 6: Distribution of final grades for the three groups of students.

Figure 6 presents the distribution of grades among the three clusters. Group 1, "The Consistents," had the best performance, with more than 70% of the students receiving an A and none of the students receiving a failing grade. Group 2, "The Sluggers," received the smallest proportion of A's. Furthermore, more than half of the students in group 2 received a D or an F. Group 3, "The Procrastinators," had high overall performance with few students receiving a failing grade. However, group 3 had less than half of the students receive an A.

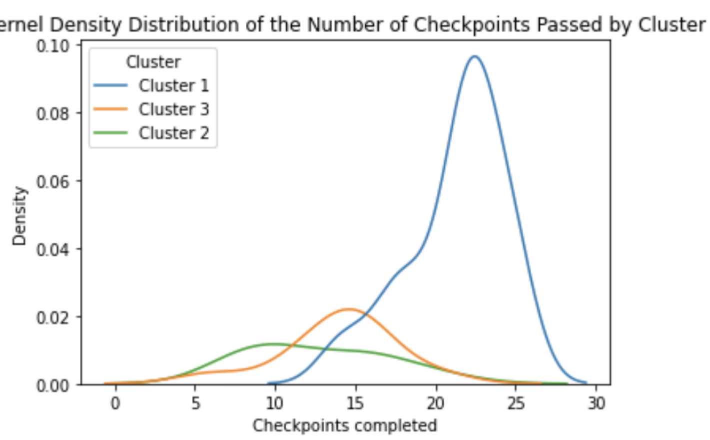


Figure 7: Distribution of final checkpoints passed between the three groups of students.

Figure 7 presents the kernel density distribution of the final number of checkpoints passed between the three groups. Students from group 1 showed the highest probability of finishing more than 22 checkpoints by the end of the semester. In contrast, students from group 2 and group 3 had a lower chance of passing more than 20 checkpoints by the end of the semester.

We also cross-matched the clustering results with the number of weekly activities students completed throughout the semester. An activity is defined to be any action or participation a student performed during the semester. For example, going to office hours, attempting a checkpoint, or revising a checkpoint. However, we found that there was no distinct pattern of activity levels between the three groups. This indicates that the lack of progress of students from group 2 can not be attributed to their lack of trying. Further analysis implies that several students from group 2 had trouble understanding the grading method. Many were able to pass most of the checkpoints, but did not meet the participation requirements to receive a passing grade.

4 Variability in Weekly Progress

We investigated whether trends in a student's weekly progress were related to their final grade in the course. Weekly progress was quantified as a student's increase in level per week. This is summarized in Figure 8. Students who achieved a D minimum final grade have fairly similar median weekly progress rates, with a slightly higher median for those who achieved an A. However, the spread in weekly progress across students for each grade are varied, especially when looking at A and B grades versus C grades.

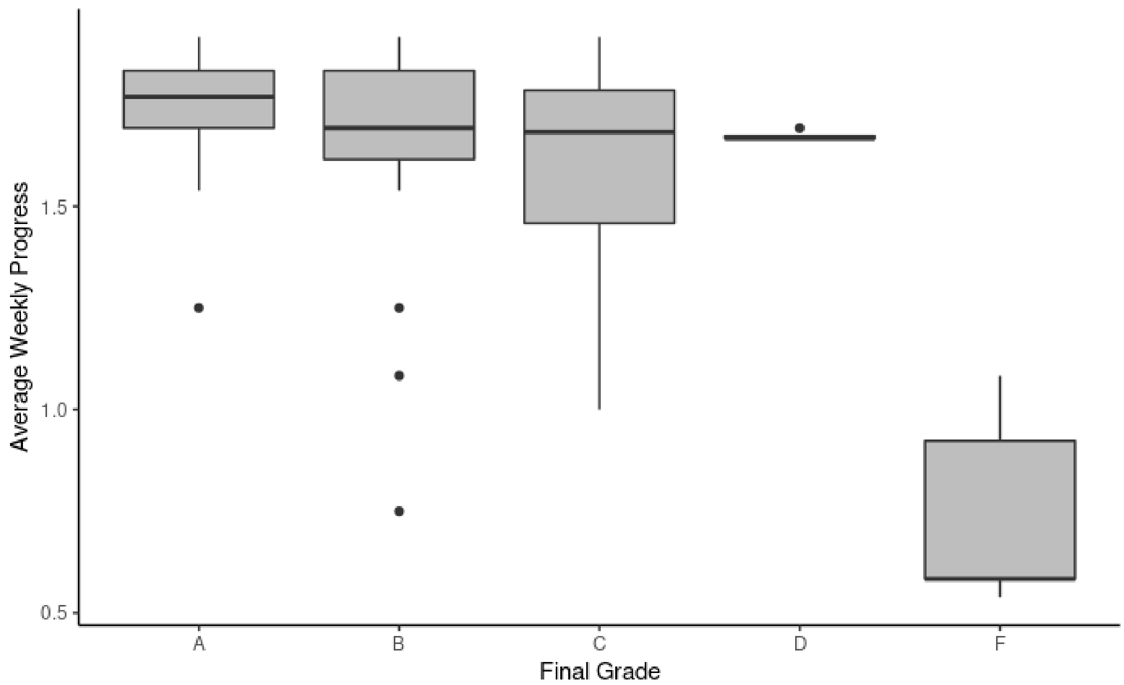


Figure 8: Relationship between average weekly progress in level and final grades for calculus I students.

To analyze this, we calculated the variance of each student's weekly progress over the semester. The variance represents how much variability exists in a student's progress from week to week. Thus, a higher variance shows inconsistent weekly progress, while a lower variance shows steady weekly progress. As shown in Table 5, Higher grades generally correlated with lower variance, while lower grades correlated with higher variance. The exception to this is those students who received an F, but the low variance for this grade does not represent steady progress but rather a lack of progress at all.

Final Grade	**Mean Progress Variance**	**Median Progress Variance**	**Standard Deviation**	**n**
A	2.295	1.889	1.323	50
B	3.824	2.806	3.034	21
C	4.209	3.899	2.682	8
D	5.816	5.243	3.487	4
F	1.602	1.234	0.861	5

Table 5: Summary of weekly progress variance of calculus I students.

Figure 9 conveys the distribution of student progress variance based on the final grade achieved. To test the significance of these trends, we used logistic regression to predict a student's final grade with their progress variance. To do this, we used binary variables to indicate whether a student achieved a C minimum, a B minimum, and an A. Logistic models were fit for each grade minimum.

Logistic regression models data using Equation (1),

$$ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x \tag{1}$$

where π is the probability of the response, x is the explanatory variable, β_0 is the y-intercept, and β_1 is the coefficient of the explanatory variable. This equation was utilized to predict three responses; final grades of a C minimum, a B minimum, and an A minimum using progress variance as the explanatory variable. We found that progress variance has little to no relationship with whether a student achieves a C minimum ($p = 0.472$), a moderate relationship with whether a student achieves a B minimum ($p = 0.093$), and is a relatively strong relationship with whether a student achieves an A ($p = 0.005$).

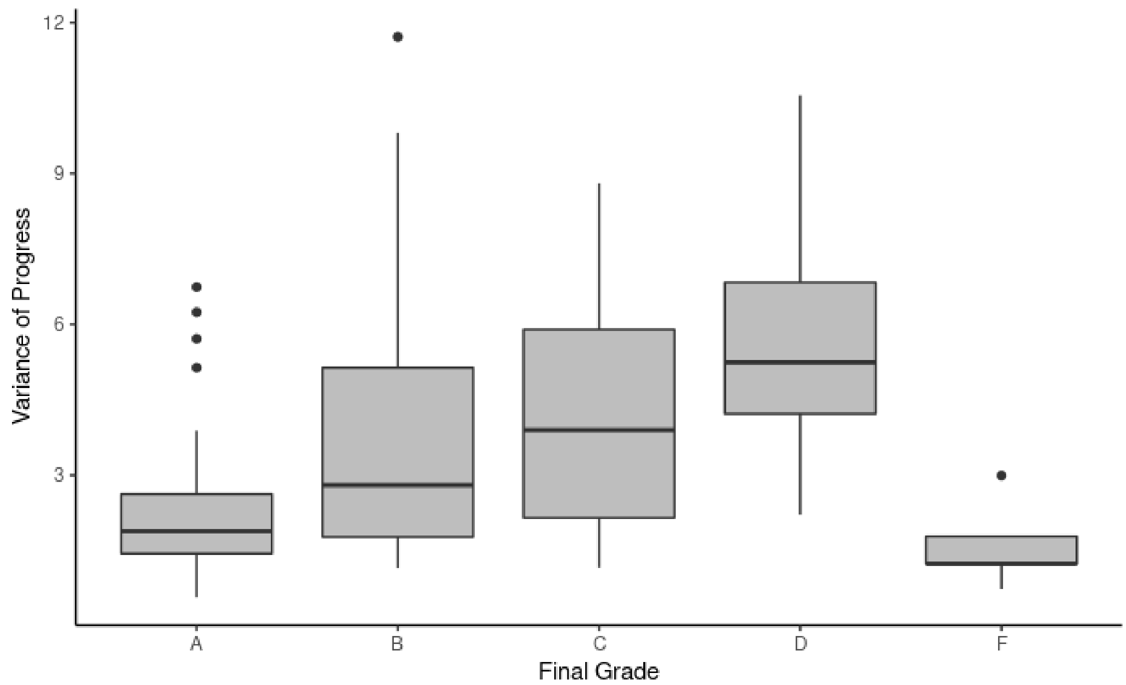


Figure 9: Relationship between weekly progress variance and final grade for calculus I students.

Students' backgrounds in mathematics were also analyzed to determine if there is a relationship to final grades. These data were collected from a pre-semester survey given to the students at the beginning of the course. Students' year in college, ratings of confidence in mathematics, and semesters since their last course in mathematics were used to predict final grades. Regardless of the minimum grade predicted, all mathematics background variables were insignificant in predicting final grades. This corresponds with the observation that there are no clear trends between final grades and these survey answers.

When considering students' average weekly progress, we found that the median values for students based on final grades were relatively close, excluding final grades of an F. However, the variability of average weekly progress for students are quite different depending on final grades. In general, we observed lower mean progress variance for students with high final grades, and higher mean progress variance for students with low final grades. The exception to this was the few students who received an F, as they exhibited an overall lack of progress. We created binary variables to indicate whether a student received a C minimum, a B minimum, or an A. Using logistic regression, we fit three models to predict these variables using progress variance. We found that progress variance is not a significant predictor of whether a student achieves a C minimum, is a moderately significant predictor of whether a student achieves a B minimum, and is a very strong predictor of whether a student achieves an A. Students' backgrounds in mathematics were not significant predictors of final grades.

5 Discussion

Using k-means clustering on student progress, we found three different patterns of behavior among students throughout the semester. Students who showed good, consistent progress were considered fully bought in and were more likely to receive better overall grades. Students who showed little activity from the beginning were considered not bought in and were far more likely to fail the class. The third pattern is where students were not considered to be bought in, but they still managed to receive decent grades because of their rapid progress near the end. This shows that, although a student may not be bought into the idea of standards-based grading, they can still work hard enough to keep their grade up.

The results from the logistic regression analysis coincided with the patterns found in the k-means clustering. Students who achieved A's showed little variability in their weekly progress, which demonstrates how these students were consistently working through the checkpoints that were required. As the grades got lower, the variability in progress increased. It was clear that students who received B's and C's had more variability when it came to level progression, which is most likely because those student struggled more with checkpoints which would lead to slower progression. An interesting result was found with students who received a failing grade. Those students also had extremely consistent level progress; however, this was because failing students showed consistently poor progress throughout the semester. Smaller variance did not necessarily mean a better grade in this case, but rather it represented a student who was consistent in their progression (or lack thereof). However, with the exception of the few students who had received F's, smaller variance in progress did correlate with higher final grades.

These results have major implications for student participation behaviors and success in standards-based grading courses. There is a significant difference in final grade distributions for students based on their patterns of behavior throughout the semester. Furthermore, students who progress more consistently throughout the semester are more likely to achieve at least a B than students who progress inconsistently. This was true regardless of students' year in college or their background and confidence in mathematics. This implies that the level to which a student buys into the standards-based grading structure of a course does correlate to how they perform in the course.

6 Limitations and Future Work

Our data were quite limited from the start. Our data were collected from only three calculus I classes at a single college, so the conclusions from the data can only be connected to one place. However, this data could be used as motivation for other professors who are also teaching standards-based courses. By encouraging their students to work consistently and not let standards pile up, the students are more likely to achieve higher grades.

Another limitation of this study was that we could not take into consideration all of the differences between students. Student attitudes toward school and learning in general can affect how they approach a course, regardless of the grading method utilized. This caused our definition of “buy-in” to have to be more generalized than we initially desired. Originally, the definition was to include how often students went to office hours and their general behavior in class. However, this idea was adjusted because outside factors such as conflicting schedules would cause students to look like they were not as bought in as others.

Future work in this area could include finding a way to implement a real-time intervention system for professors using standards-based grading. This way, they could easily track student progress and discover which cluster they are becoming a part of and be able to help students who are struggling with buying in. Future work would also include analyzing other factors that are related to student buy-in. Although we found that mathematics background was unrelated, other factors such as GPA and familiarity with standard-based grading could be related to how much students are willing to buy into a standards-based grading course.

References

[1] Freund, Rudolf J., William J. Wilson, and Ping Sa. Regression analysis. Elsevier, 2006.

[2] Hamilton, James Douglas. Time series analysis. Princeton university press, 2020.

[3] Lewis, Drew, “Student Anxiety in Standards-based Grading in Mathematics Courses” SpringerLink, 7 Dec. 2019, https://link.springer.com/article/10.1007/s10755-019-09489-3.

[4] Meyer, Luanna H., et al. “Secondary Student Motivation Orientations and Standards-Based Achievement Outcomes.” British Journal of Educational Psychology, vol. 79, no. 2, 2009, pp. 273–293., https://doi.org/10.1348/000709908x354591.