Education Influential Factors of University of Toronto Mississauga Attendance

Luai Al-Labadi[1], Hrithik Kumar Advani[2], Harsh Jhunjhunwala[3], Saarang Ahuja[4], Brittani Holder[5], Iris Wu[6], and Kyuson Lim[7]

[1,2,3,4,5,6]Department of Mathematical & Computational Sciences, University of Toronto Mississauga, Ontario L5L 1C6, Canada. E-mail: luai.allabadi@utoronto.ca, hrithik.kumaradvani@mail.utoronto.ca

[7]Department of Mathematics & Statistics, McMaster University, 1280 Main St W, Hamilton, Ontario L8S 4L8, Canada, Email: limk15@mcmaster.ca

April 16, 2022

Abstract

Absenteeism among university students is a widespread issue today. It is known that absenteeism may incur negative effects on students' academic performance as well as many social problems. This study was carried out to investigate and highlight students' perceptions of the factors affecting university attendance for online and in-person classes. The study surveyed students from a variety of disciplines at the University of Toronto Mississauga. The results of the survey indicated that the statistically significant factors affecting university attendance include gender, a student's current university CGPA (Cumulative GPA), and lectures where mandatory participation is required. Appropriate remediation to reduce the percentage of students' absenteeism are also proposed.

Keywords: COVID-19, pandemic, higher education, absenteeism, attendance, performance

1 Introduction

The term absenteeism has been defined as the conscious and deliberate act of being away from the physical space of the University classroom, conditioned for some factors that influence the search for alternatives to the use of time (Crespo et al., 2012). Within education assessment, absenteeism is a key variable that can affect students' academic performance (Hakami, 2021). Since 2019, many universities have provided online classes to prevent any hindrance to students' academic life from the COVID-19 pandemic. By doing so, students now have the liberty of completing their education at home instead of attending in-person classes. However, class attendance, whether in-person or online, is one of the most crucial factors for academic success. Attendance is a key component in students' retention, progression, achievement, and employability (Fayombo et al., 2012). The current study aims to identify the social, biological, and environmental factors that play a role in absenteeism. Moreover, this study seeks to determine possible remediations that may be imparted by university officials to increase class attendance. Some examples of the factors studied are the gender of the student, the lecture time, a lecturer's or teaching assistant's teaching method/style, as well as whether the class mandates participation through educational software. The findings of this study will incite future teaching professors, teaching assistants, and other university staff on the difficulties that students face and how these struggles may be alleviated to improve students' class attendance and performance.

2 Literature Review

The correlation between undergraduate attendance and grades has been surveyed several times before and they have all come to the same conclusion - attendance has a positive and significant effect on grades. Romer (1993), in his seminal work, found that one-third of the class was absent during a typical class meeting and noted some key points. Firstly, absenteeism seems to be lower for core courses and higher for electives. Secondly, absenteeism is lower in a smaller university as opposed to a larger one.

Romer's results have been corroborated by Schmidt (1983), Park and Kerr (1990), and Marburger (2001) who have all controlled for various factors and concluded similar results to varying degrees. While the absenteeism results have been similar in various studies, several authors mention minor inconsistencies which could be attributed to the location and the demographic. Furthermore, they note that absenteeism is higher on Fridays and seems to have an upward trend as the semester progresses forward.

When looking at the relationship of absenteeism with exam performance, we also need to understand what causes the low attendance seen in past research. The conclusions by Lotz and Lee (1999) verify the hypotheses from Marburger (2001) and Romer (1993). They find that students cite a negative self-image and low self-esteem as reasons for non-attendance. Interestingly, a view provided by Lotz and Lee (1999), Williams (1999) as well as Wadesango (2011) shows that compulsory attendance might contribute to furthering the absentee problem.

Studies indicate that absenteeism is also caused by several other factors such as lack of interesting and challenging curriculum; a desire for hedonistic activities with peers; lack of personal interest in studies; lack of confidence in a professor; inadequate relations between a student and their lecturer and commute time to university (Mayer and Mitchell 1996; Weller 1996; Williams 2000; Marburger 2001). Interestingly, Triado-Ivern et al. (2020) corroborated these findings in Barcelona and found that these factors vary based on year of education. Students in earlier years noted that their main reason for not attending the class was non-mandatory class attendance. Those in later years were more likely to attend classes based on how they managed their time.

While most of the research points to lack of interest and inadequate relations between students and the lecturer as the primary causes for absenteeism, Majeed et al. (2019) found that the main causes of absenteeism in students were Physical; where they either had health troubles, were generally too tired to attend class, or they always went to bed late which made it harder for them to attend early morning classes. Approximately 84% of the participants agreed that their cause of absenteeism was their tiredness to attending the classes.

Mohamed et al. (2018) conducted a study to determine the factors influencing student absenteeism in universities. Approximately 140 students from the Universiti Teknologi MARA, Pulau Pinang branch were surveyed. Of the 140 students surveyed, approximately 82.6% were males. It was determined that a student's attitude, activity, and family were the significant factors causing absenteeism. A student's attitude (negative academic perception, lack of interest in course material, and lack of motivation) was determined to be the dominating factor that influenced them to miss the class.

Finally, we also look at the role student gender plays on absenteeism and student CGPA. As shown in the study by Hakami (2021), gender was considered a confounding variable and he found that absenteeism is a negative predictor for males but not females studying medical sciences. These results were corroborated by Valli Jayanthi et al. (2014) who measured academic performance using CGPA and showed that student grades are affected by gender, extracurricular activities, nationality, and age.

3 Research Methodology

This study was conducted at the undergraduate sections of the University of Toronto Mississauga (UTM), Canada. A quantitative methodology was used to study the factors affecting university attendance (CGPA, commute time to the university, living in-campus/off-campus, the program of study, lecture style, and mandatory activities during lecture). Essentially, the focus is to determine whether there is any correlation between the factors mentioned above and student attendance and if so, how big is it. To answer this, the following research questions were addressed in the study:

- **RQ1.** Will a student's academic profile (a program of study, lecture style, mandatory activities, etc.) influence absenteeism?
- **RQ2.** Will a student's gender influence absenteeism? Based on practical rational and past literature, the following hypotheses were tested:
- **H1.** A student's academic profile (a program of study, lecture style, mandatory activities, etc.) influences absenteeism.
- **H2.** There will be no significant effect of gender on a student's absenteeism.

Even though the above hypothesis is based on the past literature, it should be noted that considering the pandemic, the results of this study could differ. The classes being recorded undermine the effect of student absenteeism on grades (since students can still watch the lectures while not attending them live).

The study aimed to understand the various items affecting students' attendance through observation and numerical data collection from a large sample, which will help to explain why individuals (e.g., undergraduate students) make certain decisions (Sears & Cairs, 2010). For instance, a survey design may provide quantitative descriptions regarding an individual's perception (Creswell, 2012). These aspects will be compared in consideration of both in-person and online classes.

3.1 Population

Population data for this study is from the University of Toronto Mississauga (UTM). As of 2021, UTM consists of over 15,000 undergraduate students (Male: 44.6% & Female: 55.4%; Domestic: 76.1% & International: 23.9%) and 860 graduate students (Male: 38.5% & Female: 61.5%; Domestic: 73.8% & International: 26.2%) (University of Toronto Mississauga, 2022).

3.2 Data Collection

Absenteeism data on the students were taken through a survey administered to all registered students at UTM at the end of Fall 2020. The survey was emailed to all the registered students as listed in UTM's directory (provided by the Office of the Registrar). A total of 1130 students completed it, amounting to a 7.53% return rate. The surveys were completed near the end of the fall semester in the academic year 2020-2021. The number of individuals who participated in the study is considered an appropriate sample size to represent the population (Kerjie and Morgan, 1970). All the questions asked through the survey contained categorical responses and you can refer to the survey questions in Table A1.

3.3 Survey Instrument

A quantitative survey was carried out for UTM students to analyze various factors affecting university attendance. A survey with 19 questions was designed, where each question refers to a factor that may contribute to one's attendance. The questions were designed to be as concise as possible to avoid any ambiguity. Next, the Office of the Registrar emailed the survey to all UTM registered students. The survey is anonymously completed. The final questionnaire consisted of the following items:

- Demographic information: Year of study, Degree, Gender, student status (domestic or international), and accessibility requirements.
- Academic questions: CGPA, Online/In-person classes, Lecture style, Mandatory participation during lecture (class discussions, quizzes using software such as Quercus, TopHat, etc.), and Impact of lecture attendance.
- Reasons for skipping in-person classes: Commute time to UTM, living on-campus versus off-campus, Type of commute to school (walking, bus, train, car, etc.), and Class start time.
- Reason for skipping online classes: Time of class as per the student's time zone, To study (e.g., prepare for a test or assignment), To work, Dislike of a lecturer's teaching method, Accessibility needs, Poor sleep habits, Mental health as a result of COVID-19 (stress, etc.), Technical issues (connection, equipment, etc.), General sickness (headaches, cold, fever, etc.), Easy to use external resources if needed (Internet, friends, etc.), and Others.

Before analyzing the survey's results, it was ensured that the questions in the survey measured the research topic's consistency, reliability, as well as inter-relation. To determine the reliability and consistency of the test items in the survey, Cronbach's alpha was run on the entire sample size of 1130 students. Cronbach's alpha measured the internal consistency of the scale, with preferred values between 0.7 and 1 (Reynaldo & Santos, 1999). The obtained value of Cronbach's alpha for the questionnaire is 0.72, which concludes that the items are internally consistent and hence making the questionnaire reliable.

3.4 Analytical Methods

Several Statistical models were employed to analyze the relationship between Student Attendance and the surveyed variable. First, a chi-square test for independence was performed to verify the initial hypotheses made in Section 1. The p-values obtained from the test are summarised in Section 4.2.

Second, to verify the results from the chi-squared test, a Logistic Regression Model was built where the target variable indicates whether the student identifies as someone who skips or attends the class. Label Encoding of all the categorical variables in the data was performed and then passed to the model. This first model used all the variables defined in Table A1 except one - student responses of how much they skip classes. This variable was excluded due to its relation to the dependent variable (nearly a one-to-one relation). Furthermore, we attempted to verify the results from the Logistic Regression Model by using backward stepwise regression.

Third, we performed penalized logistic regression for variable selection. These models are generally used in cases with high multicollinearity and to automate variable selection. However, while our data has low multicollinearity (or even zero), we want to verify the results of variable selection from the initial models.

L1-norm ("lasso"), L2-norm ("ridge"), and a combination of the two ("Elastic Net") models. We will discuss the results of all the models in Section 4.

These methods help us identify the existence of a correlation between absenteeism and student grades and find the statistically significant factors that affect absenteeism. Furthermore, it helps us to test any discrepancies that may arise between past literature and our paper due to the pandemic and impact of recorded lectures on attendance. However, the question that arises is of causality – Do the factors considered have a causal relation with absenteeism, or is the effect due to the presence of some exogenous variables? We will consider these in later sections.

4 Data Analysis and Results

The survey responses have been summarized and documented in the Table A2 of the appendix. Tables A3 and A4 in the Appendix look at groupings of students by gender and study year. Survey data were analyzed using R software and Python. The first part of this section describes the participant's characteristics obtained from survey data. Later, some of the findings of the survey data, established using data modeling have been discussed. Finally, this section concludes by describing the statistical significance of the results obtained from the analysis of the association among the variables of interest related to the present study's objective.

4.1 Correlation Analysis

Before diving into our analysis, we look at the correlation between the variables tested in the logistic regression model and the chi-squared test (you can find the questions corresponding to each variable in Table A1). As shown in Figure 1, all of the variables seem to be weakly correlated (or, not at all) with some exceptions which have correlations ranging from 0.15 to 0.27.

There seem to be relatively higher correlations among the lecture types. The students were asked questions on their preference of lecture delivery (traditional, hybrid, live, or recorded). Traditional and hybrid lectures have a correlation of -0.22 while traditional and live lectures have a correlation coefficient of 0.27 and recorded and hybrid lectures correlate -0.26. This helps us get a brief look at the similarity between student preferences.

Furthermore, the correlation matrix disputes some of the initial hypotheses mentioned earlier. We expected a moderate to strong negative correlation between Student Attendance and Student CGPA (Cumulative Grade Point Average) along with a relatively strong positive correlation between Student Attendance and Traditional Lectures as opposed to Hybrid Lectures. However, there is a weak correlation of -0.06 between Attendance and CGPA and no correlation between Attendance and Lecture Types.

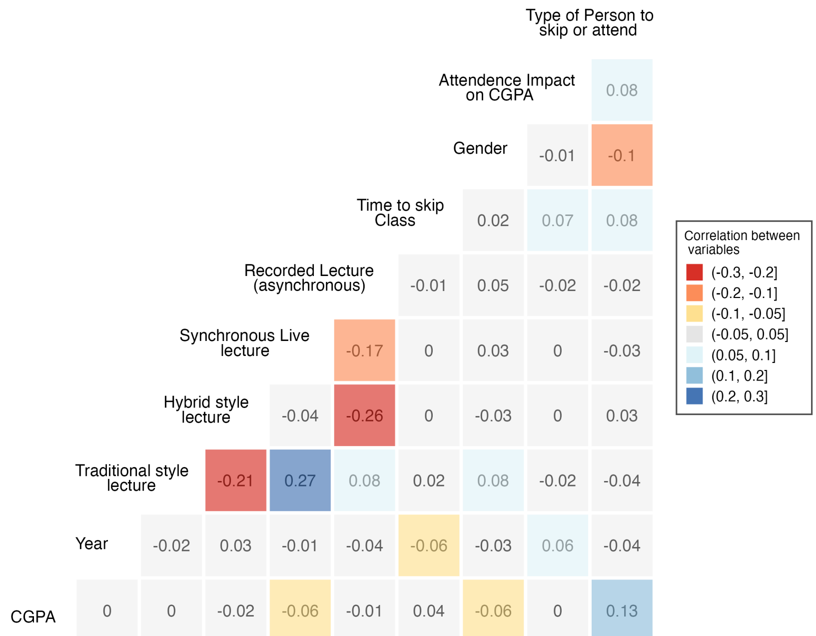

Figure 1: Correlation matrix of all variables.

While the results of the correlation analysis don't necessarily point to a relation (or lack thereof), they undermine our hypothesis in Section 3. Our initial hypothesis tested that a student's academic profile plays a role in attendance but there seems to be a weak correlation.

4.2 Characteristics of the Participants

This section looks at the survey responses and summarizes the results shown in Table A2, Table A3 and Table A4. It showcases the differences in the demographic and their replies to the questions.

4.2.1 Gender Differences

The gender distribution of the sample (1130 students) at UTM was relatively skewed with 67.9% of the respondents being females. Figure 2 (a) below, visualizes the results shown in Table A3. It shows the percentage of the class being skipped, grouped by gender, and we find that the male students are a little more likely to skip class (24.6%) as opposed to females (20%) with "Other" being the most likely to skip classes at (31.5%).

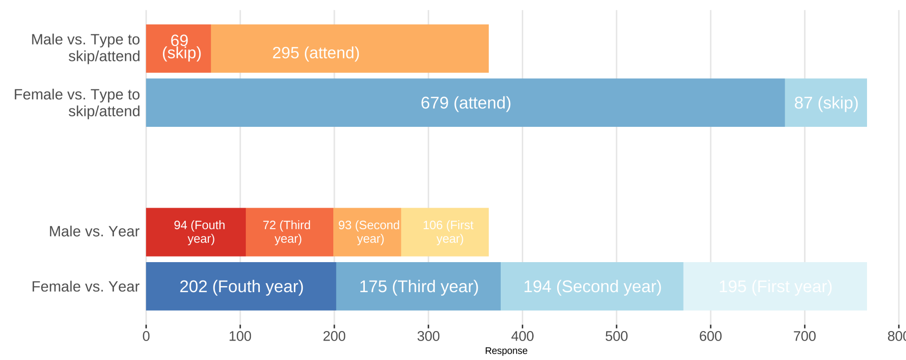

Figure 2: Stacked bar plot (a) Distribution of classes being missed, grouped by student gender. (b) Distribution of classes being missed, grouped by student gender and study year.

4.2.2 Academic Differences

As an extension of the Table A3 results, Table A4 looks at the grouped data by the year of study as well. As shown in Figure 2 (b), we notice a similar trend for all genders – the percentage of classes missed show an increasing trend in the second and third year (22% for females and 30% for males) and are lower in the first and fourth year (15% for females and 20% for males). While there was similar participation from the students studying in different years of the university, 78.9% of the students were domestic and 57.3% of them were pursuing an Honors Bachelor of Science degree. Most students ($n = 1067, 96.1\%$) were taking only online classes. This was mainly because the survey was conducted during the COVID-19 pandemic. There was a mixed response as to what kind of lectures (In-Person, Live, Recorded) the students would prefer. It was noted that the most frequently missed classes by the students ($n = 646, 58.2\%$) were the morning classes occurring between 7 am and 12 pm (Eastern Standard Time). Overall, most students ($n = 936, 84.3\%$) indicated that they would not skip a class where attendance in a lecture was mandatory, which included participating in class discussions and quizzes.

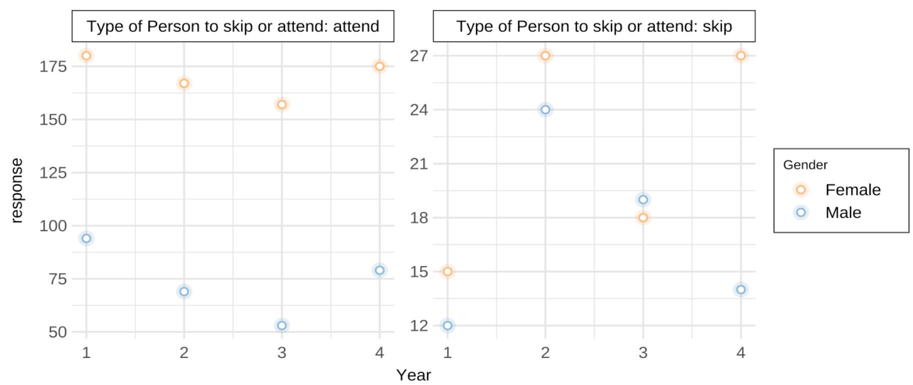

Figure 3: Point plot (a) Distribution of classes being missed, grouped by student gender. (b) Distribution of classes being missed, grouped by student gender and study year.

4.2.3 Student Perceptions and Replies

Most of the students ($n = 742, 66.8\%$) strongly agreed that attendance affects their overall CGPA. Most of the students ($n = 954, 85.9\%$) indicated they attended their classes. Approximately 37% of the students indicated that they would rarely ($< 5\%$) skip a class. Moreover, the top three reasons indicated by students as to why they miss their classes were studying for the test or assignment during class time ($n = 668, 60.1\%$), mental health issues due to COVID-19 ($n = 589, 53.1\%$), and poor sleeping habits ($n = 461, 41.5\%$). This information is displayed in Table A1 of the appendix.

4.3 Statistical Tests and Models

An investigation into how the attendance of students varies with variables of interest related to the present study's objective was also conducted. The contingency tables (crosstabs) between attendance and other factors can be found in Appendix Table A5. Additionally, in the analysis below we have referred to the effect of a variable called "Attendance impact on CGPA". This variable identifies the student's replies to the question "Do you think Attendance impacts your CGPA?" and is different from the 'actual' impact of Attendance on student CGPA. The 'actual' relation between these two variables is not being tested in this study. However, we do expect differences from the past literature, owing to the pandemic.

4.3.1 Chi-Square Test

To investigate which of the factors in the questionnaire vary by attendance, a chi-squared test was conducted, which in turn helps us answer the research questions we asked earlier. The p-value of all the variables is summarized in Table 1 below. It was found that there are four statistically significant tests (p-value < 0.05). Those significant tests are for testing whether there is a relationship between Attendance and four other factors; Gender, Degree type, CGPA, and Attendance in lectures where mandatory participation is required. There is enough evidence to conclude that these factors vary by attendance. Using the chi-squared test, we can accept both the hypotheses that we built in earlier sections.

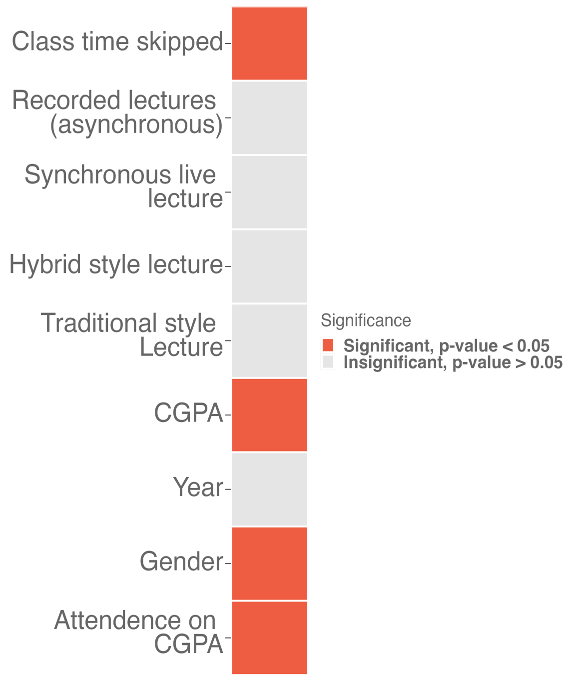

Figure 4: A p-values of chi-square test of independence between attendance and the variables of interest.

4.3.2 Logistic Regression Model

An alternate method was employed to check the consistency of our results. As mentioned earlier, the data consisted of only categorical variables and that is why the first model we chose was a Logistic Regression Model with a dichotomous dependent variable - "attend miss class". This variable indicates whether the student identifies as someone who skips or attends the class.

It is evident by the results in Table 2 that the feature "Mandatory Graded Assignments/Quizzes During Class", is the most statistically significant factor with a p-value of 0.0000143. "Mandatory Graded Assignments/Quizzes During Class" refers to students being graded on assigned assignments or quizzes during class, usually through software such as ’TopHat’. In Table 2, depending on what the p-value is, the significance column ranges from ’*’ to ’***’, with ’*’ representing low statistical significance (p-value less than 0.05) and ’***’ the highest statistical significance (p-value near 0.00). This implies that students are more likely to attend classes if the professor is to include in-person graded quizzes or activities during lecture time. Additionally, ”Student CGPA”, ”Attendance Impact on CGPA”, ”Student Gender” and ”Timing of Missed Classes” appear as relatively significant factors with p-values of 0.000847, 0.010975, 0.046513, and 0.0000059, respectively. These results were found to be consistent with the previous chi-square test findings.

Furthermore, the summary results of the Logistic Regression model show the Null Deviance, Residual Deviance, and the AIC. Using the Null and Residual Deviance, we find that our χ^2 value with 14 degrees of freedom gives us a p-value of nearly 0. Thus, we can conclude that the model is highly useful for predicting the factors of absenteeism. Also, the AIC value of this initial model is 860.98. Akaike Information Criterion (AIC) helps understand the relative quality of statistical models for a given dataset.

$$Y_i = \beta_0 + \beta_1 CGPA_i + \beta_2 \text{attendence impact}_i + \cdots$$

$$+\beta_3 \text{ Study year}_i + \beta_{12} \text{ recorded lecture}_i$$

$$+\beta_{13} \text{ mandatory attendence}_i + \epsilon_i$$

It is evident by the results in Table 2 that the feature ”Mandatory Graded Assignments/Quizzes During Class”, is the most statistically significant factor with a p-value of 0.0000143. ”Mandatory Graded Assignments/Quizzes During Class” refers to students being graded on assigned assignments or quizzes during class, usually through software such as ’TopHat’. In Table 2, depending on what the p-value is, the significance column ranges from '*' to '***', with '*' representing low statistical significance (p-value less than 0.05) and '***' the highest statistical significance (p-value near 0.00). This implies that students are more likely to attend classes if the professor is to include in-person graded quizzes or activities during lecture time. Additionally, "Student CGPA", "Attendance Impact on CGPA", "Student Gender" and "Timing of Missed Classes" appear as relatively significant factors with p-values of 0.000847, 0.010975, 0.046513, and 0.0000059, respectively. These results were found to be consistent with the previous chi-square test findings.

Furthermore, the summary results of the Logistic Regression model show the Null Deviance, Residual Deviance, and the AIC. Using the Null and Residual Deviance, we find that our χ^2 value with 14 degrees of freedom gives us a p-value of nearly 0. Thus, we can conclude that the model is highly useful for predicting the factors of absenteeism. Also, the AIC value of this initial model is 860.98. Akaike Information Criterion (AIC) helps understand the relative quality of statistical models for a given dataset.

Question	Coefficient	
1. CGPA		
about the impact of COVID-19 on covering your tuition cost	0.666	*
1-2. For the current semester, how concerned are you about the impact of COVID-19 on using up your saving	0.676	*
1-3. For the current semester, how concerned are you about the impact of COVID-19 on your increased debt	0.817	*
1-4. For the next semester, how concerned are you about the impact of COVID-19 on covering your tuition cost	0.676	*
1-5. For the next semester, how concerned are you about the impact of COVID-19 on using up your saving	0.727	*
1-6. For the next semester, how concerned are you about the impact of COVID-19 on your increased debt	0.843	*
1-7. How concerned are you about the impact of COVID-19 on your ability to meet your overall financial obligation or essential needs	0.720	*

Table 2. A 2-factor analysis model by principal component method

4.3.3 Backward Stepwise Regression

Following the Logistic Regression model, we attempted to verify the results through Stepwise Regression Models. First, we performed a Backward Stepwise Regression with the results shown in Table 3.

The backward stepwise regression helps us narrow down the important variables. It retained all the statistically significant variables from the Logistic Regression along with the intercept, "Study Year" and the "Traditional Lecture Rating". The latter variable shows the preferences students have for traditional (in-person) lectures. The deviance results showed that the Stepwise Regression was useful as well and the AIC for this model was 851.28.

Question	Coefficient	
1. CGPA about the impact of COVID-19 on covering your tuition cost	0.666	*
1-2. For the current semester, how concerned are you about the impact of COVID-19 on using up your saving	0.676	*
1-3. For the current semester, how concerned are you about the impact of COVID-19 on your increased debt	0.817	*
1-4. For the next semester, how concerned are you about the impact of COVID-19 on covering your tuition cost	0.676	*
1-5. For the next semester, how concerned are you about the impact of COVID-19 on using up your saving	0.727	*
1-6. For the next semester, how concerned are you about the impact of COVID-19 on your increased debt	0.843	*
1-7. How concerned are you about the impact of COVID-19 on your ability to meet your overall financial obligation or essential needs	0.720	*

Table 2. A 2-factor analysis model by principal component method

4.3.4 Penalized Logistic Regression

Once we had agreeable results from the two Logistic Regression models along with the chi-squared test, we ran a final model for variable selection. For this third model, we used penalized logistic regression which has three categories - L1-norm (”lasso”), L2-norm (”ridge”), and a combination of the two (”Elastic Net”). Lasso and Ridge are extensions of the original regression with penalty terms added. The treatment of the penalties in these models is the only difference between them. Furthermore, the Elastic Net approach is a combination of both the penalty terms with α signifying the differentiator. If α is set to 0, we get a ridge regression and if it's set to 1, we get a lasso regression. For an elastic net, α can be any value between 0 and 1 (depending on which penalty we want to prioritize more). In our model, we have used an α of 0.5.

$$Y_{\text{Lasso}} = \arg\min_{\beta}(||Y - \beta \times X||^2 + \lambda||\beta||_1)$$

$$Y_{\text{Ridge}} = \arg\min_{\beta}(||Y - \beta \times X||^2 + \lambda||\beta||_2^2)$$

In both the equations above, the first part corresponds to the Residual Sum of Squares (RSS) and the second part corresponds to the penalty applied. A β is the magnitude of coefficients and λ is the amount of shrinkage. The λ value helps control for excess variables in the model. Table 4 shows the results of the coefficients of all three penalized logistic regression models, while Figure 3 visualizes these same results. Also, Figure 4 shows the cross-validation results of the models to find the best lambda. Table 4 shows us that the penalized regression models agree with the logistic regression models we built earlier. The coefficients of Lasso, Ridge and the Elastic Net, all show the same variables as significant. However, in the case of Ridge Regression, it finds that "Class Type" also has some significance in predicting Attendance. This effect has not been seen in any other models.

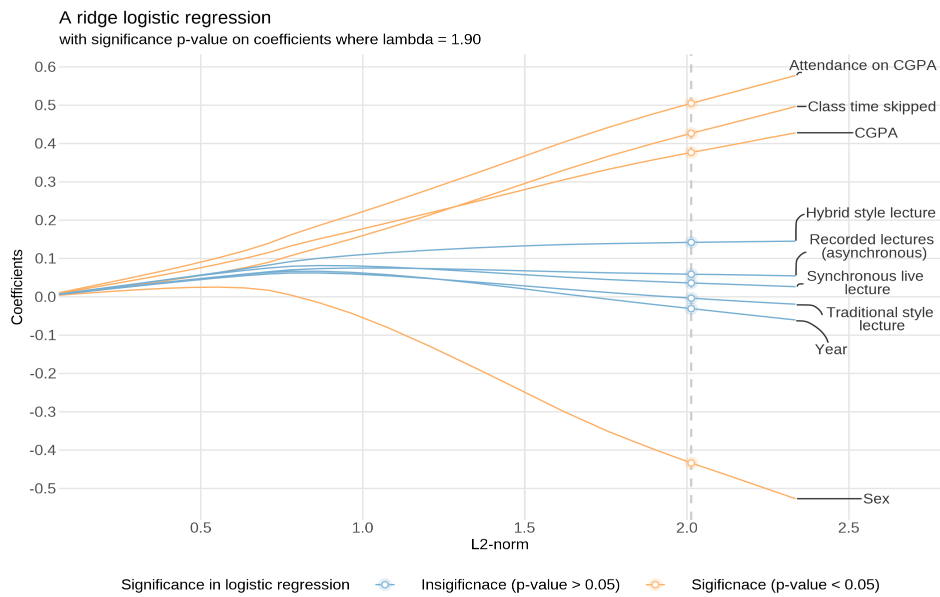

Figure 5: Coefficient results of the penalized logistic regression models.

4.3.5 Limitations of the Models

We have considered several models in our analysis but it's important to keep a note of the limitations they present and any cases we may not have accounted for. Firstly, the lack of interaction effects. From our initial analysis, we tested some interaction effects between predictor variables and found them to present in some cases. However, we chose not to include them since these variables were also highly correlated and the interaction effect require further

5 Discussion and Proposed Solution Strategies

Throughout the paper, we have looked at the results from the survey conducted at the University of Toronto Mississauga and performed several tests. The results from our initial analysis, along with the chi-squared test and the Logistic Regression seem to agree not only amongst each other but also with the past literature. Statistical analysis of the data suggests that the statistically significant factors associated with university attendance are gender, degree type, current CGPA (cumulative GPA), and lectures where mandatory participation is required (class discussions, quizzes, etc.). When considering the categorical effect of gender, we find that male students are significantly more absent than female students and these absenteeism results are corroborated by past literature in Nja et al. (2019) and Khan (2018).

This paper also looks at the CGPA of students as a determinant of absenteeism and the results are consistent with those found in the past literature. Researchers have emphasized the importance of lecture attendance for one's academic success (or, CGPA). Specifically, the studies completed by Aden et al. (2013), Kassarnig et al. (2017), and Nja et al. (2019) reflect these sentiments. Finally, the most significant factor affecting attendance was found to be the lecture delivery. Lectures, where mandatory participation is required, had the highest attendance.

When looking at potential solutions to resolving the absenteeism problem, the solutions seem to depend on each other. When addressing absenteeism in males, one approach may be to encourage male students to attend classes by providing counseling sessions for students to highlight the importance of attending classes, and its consequent impact on their performance henceforth. Therefore, mandatory participation in classes may be the key to resolving the problem of absenteeism. This may be implemented in a variety of ways. Instructors may increase the number of in-class quizzes and participation while reducing the weight given to term-end exams. Instructors may also utilize technology in class such as Top Hat, iClicker, and Kahoot, which will optimize students' attendance while identifying the student's level of understanding, allowing for improvement earlier on in the course (Al-Labadi and Sant, 2021).

This paper also looks at the student's reasons for absenteeism during lecture time. The students gave several reasons for this, but the most common reason given for skipping online classes was to study and prepare for the other courses during the lecture time (60.1% of respondents). This includes preparing for the assignments, quizzes, or tests for other courses that they were taking during the term. Similar studies like the one conducted by Kottasz (2005) support these results, indicating this factor as one of the key reasons leading to student absenteeism. To remediate this, the university may perhaps organize work/time management workshops, as well as providing one on one advice sessions to students. Alternatively, the university could help the student build a better course timetable at the start of the semester to avoid running into these problems.

The second reason selected by students as to why they may miss an online class was their current mental health situation (53.1% of respondents). Sometimes students may have to face personal or family-related issues, and absence in class is unavoidable following these circumstances. The stress and pressure that arise from these situations tend to have a major impact on a student's overall well-being. When students face conflicts from school or home, this may result in a reduction of focus during class which in turn affects their academic performance as well. To ensure students have access to any support they may need during this time, it is recommended to encourage students to be aware of the resources provided by the university and provide one on one counseling sessions for students going through the difficult phase.

Another reason indicated by students for skipping the class was poor sleeping habits. According to the findings, 58.2% of the students responded that they are more likely to miss classes in the morning than at other times of the day. Thus, the study found that there was a negative relationship between the quality of sleep and the absenteeism rate for students. A likely cause for this behavior is students staying up late at night, which may result in their failure to attend early morning classes at the scheduled time. It is theorized that students who have a lack of sleep are more likely to be absent because of fatigue associated with staying up late to work or study in addition to the long commute time between school and home. Some approaches to increasing students' attendance may be by providing more class times in the afternoon or evening, as well as increasing bus routes for students who need to travel long distances to attend in-person classes. An alternative approach under the control of the university could be to work with the student to build a more suitable timetable at the start of the semester.

6 Limitations and Conclusion

The current study was conducted during the COVID-19 pandemic; therefore, the majority of the responses were reflective of students currently attending classes online. It would be interesting to investigate whether the same factors would be selected as the reason for absenteeism when in-person classes resume. This in turn points to the differences with past literature. With several classes being recorded, students can maintain similar grades while missing more classes thereby reducing the correlation that has been seen between these variables.

Secondly, this survey was limited to students from the University of Toronto Mississauga (UTM), therefore the sample size with a variety of degrees were limited as UTM currently offers 4 types of degrees (Honours Bachelor of Arts, Honours Bachelor of Science, Bachelor of Commerce, and Bachelor of Business Administration). Including students from other campuses of the University of Toronto may be considered for future work which will incorporate students from other degree types, giving us more in-depth information about the factors causing absenteeism among students. Furthermore, including students from other campuses might help to tap into a new demographic of students whose commute times differ, and residence locations differ.

However, we do have to consider the limitations of including more institutions as well. Different campuses or universities may lead to differences in course structure and delivery because of which the survey and the experiment might have to be edited. It might be more appropriate to use Stratified Random Sampling instead of Simple Random Sampling as in this paper.

Furthermore, a possible limitation is the sample selection. From the 1000 respondents, over 50% of the students replied that they would miss a class less than 5% of the time. There is a potential bias since this was an optional survey and students who attend class or do well in class might be the ones who participate.

Finally, the design of the questionnaire could have also hindered the results. It is possible that other important factors were missed due to the survey preselected factors and not allowing for an open-ended question/answer response. The curation and selection of the factors subsequently resulted in a more controlled study, of which there is a possibility that some other important factors were missed.

The results of this study indicate that the factors ultimately affecting university attendance are the student's gender, current CGPA (Cumulative GPA), and lectures where mandatory participation is required. Universities may help by providing support and counselling services to students who may be struggling personally, which inadvertently affects university attendance. The university may also provide more options for students by providing more class time options to attend class, as well as utilizing different technologies in the classroom to increase engagement and university attendance overall.

Past literature has already shown the importance of absenteeism on academic performance and this paper helps drive that point home while showing the significant factors that affect a student's absenteeism. We hope that this paper would lead to new research in this area to account for the limitations and that it would encourage universities to work with students and help reduce absenteeism. It's not a problem that students face alone, and we have seen that the primary reasons for absenteeism are a busy schedule and mental health – both of which could ideally be curbed with some help.

7 Ethical Statement

Comprehensive from the ethics standard at University of Toronto Mississauga. This has been authorized based on the approved number: 2019-036 from the ethics review committee by the University of Toronto.

References

[1] Aden, A., Yahye, Z., & Dahir, A. (2013). The effect of student's attendance on academic performance: a case study at Simad University Mogadishu. Academic Research International, 6(4), 409-417.

[2] Agresti, A. (2018). Statistical Methods for the Social Sciences, 5th edition. Boston: Pearson Education, Inc.

[3] Al-Labadi, L., & Sant, S. (2021). Enhance learning experience using technology in class. JOTSE: Journal of Technology and Science Education, 11(1), 44-52.

[4] Cecilia, N., Bernedette, C., & Christiana Chinyere, I. (2019). The influence of age and gender on class attendance plus the academic achievement of undergraduate chemistry education students at University of Calabar. Educational Research and Reviews, 14(18), 661-667.

[5] Creswell, J. W. (2012). Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research. Boston, MA: Pearson Education, Inc.

[6] Jayanthi, S. V., Balakrishnan, S., Lim, A., Abdul Latiff, N. A., & Nasirudeen, A. M. A. (2014). Factors Contributing to Academic Performance of Students in a Tertiary Institution in Singapore. American Journal of Educational Research, 2(9), 752-758. doi:10.12691/education-2-9-8

[7] Kassarnig, V., Bjerre-Nielsen, A., Mones. E., Lehmann, S., & Lassen, D. D. (2017). Class attendance, peer similarity, and academic performance in a large field study. PLOS ONE, 12(11): e0187078. https://doi.org/10.1371/journal.pone.0187078

[8] Khan, S. N. (2018). Attendance at university lectures: A study into factors which influence it, and exam performance as a consequence of it. MSOR Connections Journals, 17(1), 46-52. https://journals.gre.ac.uk/index.php/msor/article/view/902

[9] Kirby, A., & McElroy, B. (2003). The effect of attendance on grade for first-year economics students in University College Cork. The Economic and Social Review, 34(3), 311-326.

[10] Kottasz, R. (2005). Reasons for student non-attendance at lectures and tutorials: an analysis. Investigations in university teaching and learning, 2(2), 5-16.

[11] Krejcie, R., & Morgan, D. (1970). Determining sample size for research activities. Educational and Psychological Measurement, 30, 607-10.

[12] Majeed, A., Hussain, M., Afzal, M., & Gilani, S. A. (2019). Causes of Absenteeism among University Students. European Academic Research. Vol.VII, Issue 3. https://www.euacademic.org/UploadArticle/4031.pdf

[13] Marburger, D. R. (2001). Absenteeism and undergraduate exam performance. The Journal of Economic Education, 32(2), 99-109. DOI: 10.1080/00220480109595176

[14] Mohamed, S. A., Ul-Saufie, A. Z., Ahmad, H., & Zahari, M. F. (2018). Factors influencing student absenteeism in the university. AIP Conference Proceedings 1974, 050013 (2018); https://doi.org/10.1063/1.5041713

[15] Nja, C., Cornelius-Ukpepi, B., & Ihejiamaizu, C. C. (2019). The influence of age and gender on class attendance plus the academic achievement of undergraduate chemistry education students at University of Calabar. Educational Research and Reviews, 14(18), 661-667

[16] Reynaldo, J. A., & Santos, A. (1999). Cronbach's alpha: a tool for assessing the reliability of scales. Journal of Extension, 37(2), 1-4.

[17] Sears, A., & Cairns, J. (2010). A Good Book, in Theory: Making Sense through Inquiry, 2nd edition. Toronto: University of Toronto Press.

[18] Triado-Ivern,X. Aparicio-Chueca, P. Elasri-Ejjaberi, A., Maestro-Yarza, I., Bernardo, M., & Presas Maynegre, P. (2020). A factorial structure of university absenteeism in higher education: A student perspective. Innovations in Education and Teaching International, 57(2), 136-147. DOI: 10.1080/14703297.2018.1538896

[19] University of Toronto Mississauga, https://www.utoronto.ca/university-life/campuses/mississauga. Retrieved January 11th, 2022.

[20] Wadesango, N., & Machingambi, S. (2011). causes and structural effects of student absenteeism: a case study of three South African universities. Journal of Social Sciences, 26(2), 89-97. DOI: 10.1080/09718923.2011.11892885

8 Acknowledgment

The authors are grateful to the Associate Editor and the reviewer for their comments and suggestions that have greatly improved the paper.

9 Appendix

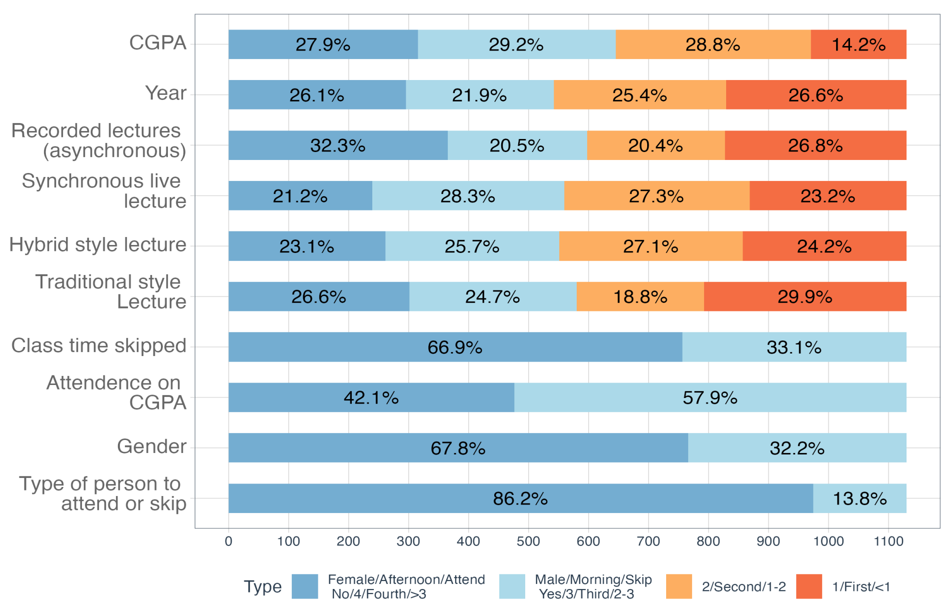

Figure 6: Likert type graph: Overall ratings of 1100 students in attendence