

**Does academic probation keep students in college or push them out?**

**The factors that lead to an individual being put on academic probation and how that affects their academic success.**

**Introduction**

About 20<sup>1</sup> percent of all college students find themselves on academic probation after their first year of college. For a student to be considered for academic probation, their GPA must be below a specified cutoff value set by their campus. It is at that point that they are sent a letter detailing the reason for their probation and outlining the conditions that they must meet in order to regain good academic standing. These conditions include raising their cumulative GPA to an acceptable level. A student who is unable to meet these conditions then runs the risk of being suspended for a full academic year. Currently, there is a debate surrounded around whether academic probation keeps students in college or pushes them out. This paper will attempt to investigate the effects that being placed on academic probation has on academic performance while also seeking to uncover the factors that lead to a student attaining poor academic standing.

**Data**

The data used in this report comes from a cross-sectional dataset of students from a large university in Canada. The university is split into 3 campuses. One central campus (*campus1*) with a GPA cutoff of 1.5 and two smaller satellite campuses (*campus2* and *campus3*) with a cutoff of 1.6. Each observation in the dataset represents a student randomly sampled from the university. Tables 1, 2, and 3 show the variables in the dataset.

**Table 1**

Categorical Variable	Description	Dummy Variables	Outcome	Mean
sex	Gender of the student.	female	=1 if female	.6260
		male	=1 if male.	.3740
mtongue	Native language spoken by the student.	english	=1 if spoken language is English	.7171
		french	=1 if spoken language is French	.0050
		other_lang	=1 if spoken language is neither English nor French	.2779
birthplace	Where the student was born.	america	=1 if the student is from America	.0019
		asia	=1 if the student is from Asia	.0770
		canada	=1 if the student is from Canada	.8784
		other	=1 if the student is from neither America, Asia, nor Canada	.0428

**Table 2**

Binary Variable	Description	Outcome	Mean	Standard Deviation
gradin4	If the student graduated in 4 years.	=1 if the student graduated in 4 years	.5444	.4980
probation_year1	If the student was placed on probation after the first year.	=1 if the student was placed on probation after the first year	.1142	.3180
probation_year2	If the student was placed on probation after the second year.	=1 if the student was placed on probation after the second year	.0404	.1968
probation_ever	If the student was ever placed on probation through their first 2 years of college.	=1 if the student was ever placed on probation in college	.1333	.3399
campus1	If the student attended campus 1.	=1 if the student attended campus 1	.6084	.4881
campus2	If the student attended campus 2.	=1 if the student attended campus 2	.1781	.3826
campus3	If the student attended campus 3.	=1 if the student attended campus 3	.2135	.4098

**Table 3**

Continuous Variable	Description	Mean	Standard Deviation	Min	Max
GPA_year1	First year GPA of the student.	2.5432	.8228	0	4.3
GPA_year2	Second year GPA of the student.	2.5874	.8138	0	4.3
GPA_total	The GPA of the student through their first 2 years	2.5653	.7542	0	4.3
totcredits_year1	Total credits attempted year 1.	4.6074	.4930	3	6
totcredits_year2	Total credits attempted year 2.	4.3426	.8958	0	9
age_at_entry	Age of the student when they entered the university.	18.6350	.7348	17	21
gpacutoff	The cut-off gpa for the student.	1.5213	.0410	1.5	1.6
hsgrade_pct	High school grade percentile ranking	52.3431	28.5243	1	100

### Regressions and Methodology

All regressions were conducted in accordance with the OLS assumptions. Meaning that, unless otherwise noted, the following regressions were created in such a way so that they were unbiased and consistent. All models were linear in relationship with data acquired from a randomly sampled population. Additionally, in such cases where homoscedasticity of the error terms was not possible, the necessary procedures were taken in order to account for their heteroskedasticity. While there were key variables of interest in each regression, many of the other variables were also included as to ensure the zero conditional mean assumption would hold. More detail will be provided when discussing the methodology.

### Factors Leading to Probation

In order to test for the factors that lead to a student being placed on academic probation, the variable *probation\_ever* was created. It was created in order to understand the total effect that

the regressors might have on probation. It takes on the value 1 if a student was placed on probation in either of their 2 years at college. The idea behind the variable was that the model could estimate the effects of why the student went on probation in either year rather than in just one. That way, a more complete sense of the effects could be obtained. The initial linear probability model (LPM) was as shown:

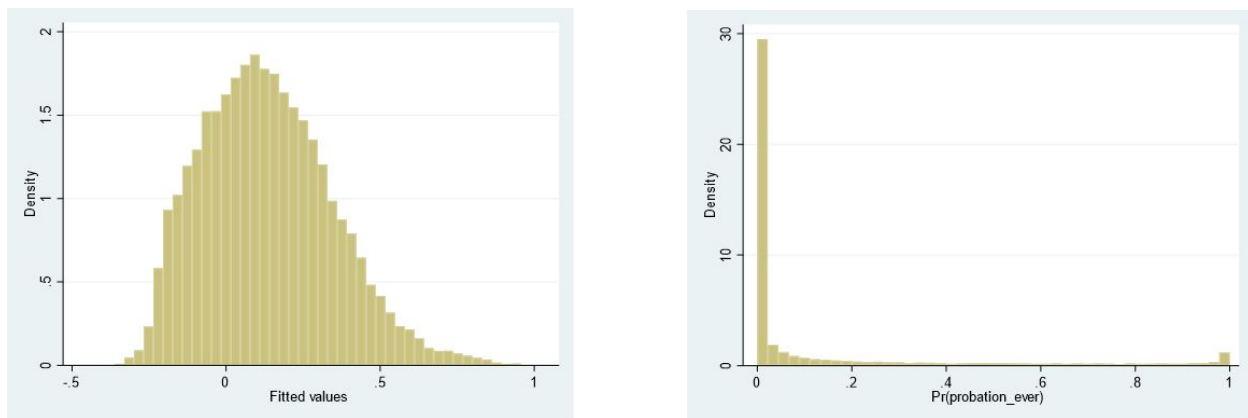
$$\begin{aligned} \text{probation\_ever} = & \beta_0 + \beta_1 \text{GPA\_total} + \beta_2 \text{totcredits\_year1} + \beta_3 \text{totcredits\_year2} + \beta_4 \text{age\_at\_entry} \\ & + \beta_5 \text{gpacutoff} + \beta_6 \text{campus2} + \beta_7 \text{hsgrade\_pct} + \beta_8 \text{female} + \beta_9 \text{english} + \beta_{10} \text{french} + \beta_{11} \text{america} + \\ & \beta_{12} \text{asia} + \beta_{13} \text{canada} + u \end{aligned}$$

In this model, every explanatory variable that could have an effect on probation was regressed. To narrow down the model, the variables were tested at a 5% significance level. The regressors with p-values greater than 0.05 could not be proven to have partial effects that were statistically significant from 0, and were thrown out of the model. The variables with p-values less than 0.05 were kept. The final model is as shown:

$$\begin{aligned} \text{probation\_ever} = & \beta_0 + \beta_1 \text{GPA\_total} + \beta_2 \text{totcredits\_year1} + \beta_3 \text{totcredits\_year2} + \beta_4 \text{gpacutoff} + \\ & \beta_5 \text{campus2} + \beta_6 \text{hsgrade\_pct} + u \end{aligned}$$

The variables *campus3* and *campus1* had to be omitted from the regression in order to avoid perfect collinearity with *gpacutoff*. Had they been left in, then the regression would have become biased. As *gpacutoff* is likely highly correlated with probation, it was a variable of key importance to test. This factored into the decision of which variables to omit. Seeing as students often list taking too many credits<sup>1</sup> as one of the main reasons for their academic deficiencies, *totcredits\_year1* and *totcredits\_year2* were additional variables of interest. Due to this being an LPM there were a few roadblocks in the regression process. Namely, that the error terms were

heteroskedastic and that the probabilities fell out of the range [0,1]. To fix the heteroskedasticity of the errors, robust standard errors were used. To allow the fitted values to fit within the required range a probit model was created. Below shows the histograms for the fitted values both before and after the probit model:



However, as you can see, STATA was unable to completely converge the fitted values. This was due to GPA being a near perfect predictor of probation. Coefficients from the model were able to be obtained. Because probit models' coefficients are difficult to interpret, a partial derivative was needed. Using the margins command, a usable partial effect was obtained for all of the variables. The margins command calculates the average marginal effect that each regressor has across all individuals. This gives the best estimate as to how the probability of probation changes given the independent regressors. The regression results can be seen in table 4:

**Table 4: *Probation\_ever* Margins**

Explanatory Variable	dy/dx	Std. Error	P >  z	95% Confidence Interval	
GPA_total	-.2441348	.0019402	0.000	-.2479375	-.240332
totcredits_year1	-.0084179	.0022766	0.000	-.0128799	-.0039559
totcredits_year2	-.0082925	.0012427	0.000	-.0107282	-.0058568
gpacutoff	.1564069	.0286546	0.000	.1002449	.2125688

campus2	.0067873	.0031097	0.029	.0006925	.0128822
hsgrade_pct	-.0003189	.0000533	0.000	-.0004234	-.0002143

Most of the variables were highly significant as they had p-values of 0. GPA and GPA cutoff were proven to have the largest effects on probation. Holding all other variables constant, when a student's GPA increases by 1 point they are predicted to be 24.4% less likely to be placed on academic probation. This result makes sense as a student who has a higher GPA is unlikely to drop below the GPA cutoff for probation. Speaking of GPA cutoff, when the threshold for academic probation is increased by 1 grade point, students are predicted to be 15.6% more likely to get placed on probation. Again, this result makes sense as when the cutoff line is higher that means that students have less opportunity for error. As theorized earlier, the amount of credits also plays an effect on the outcome of probation, though it is not the expected outcome. As it turns out, increasing the amount of credits taken in either year of college leads to a student being less likely to be placed on probation. Students are .8% less likely to go on probation when they take 1 additional credit in either year. While this change is small, it is still interesting as it contradicts what many believe to be one of the causes behind student deficiencies.

### Long Term Academic Success

To estimate the effects that probation has on academic success, the dependent variable *gradin4* was used. Most college programs are designed to be completed in 4 years and, as such, a student who has completed their school work within 4 years is considered to have been on time. For a student to have been on time they must have put in the work and passed their classes.

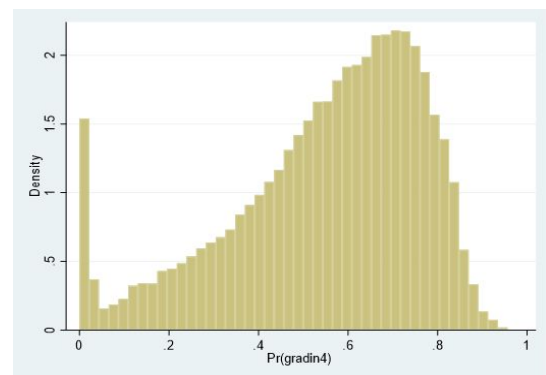
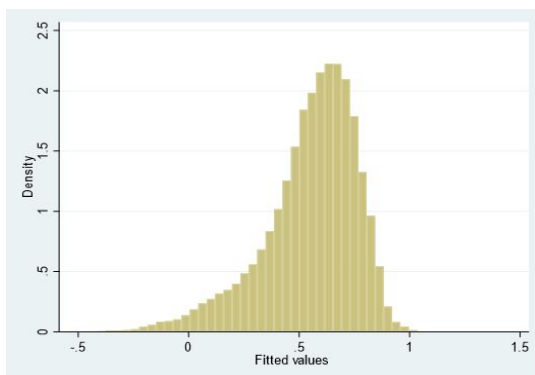
Therefore, graduating in 4 years is a good indication of academic performance throughout college. Similarly to the previous regression, the dependent variable (*gradin4*) was regressed across several other potentially correlated regressors and tested at the 5% significance level. The initial model is as shown:

$$\begin{aligned} \text{Gradin4} = & \beta_0 + \beta_1 \text{GPA\_year1} + \beta_2 \text{GPA\_year2} + \beta_3 \text{totcredits\_year1} + \beta_4 \text{totcredits\_year2} + \\ & \beta_5 \text{age\_at\_entry} + \beta_6 \text{probation\_ever} + \beta_7 \text{suspended\_ever} + \beta_8 \text{gpacutoff} + \beta_9 \text{campus2} + \\ & \beta_{10} \text{hs\_grade\_pct} + \beta_{11} \text{female} + \beta_{12} \text{english} + \beta_{13} \text{french} + \beta_{14} \text{america} + \beta_{15} \text{asia} + \beta_{16} \text{canada} + u \end{aligned}$$

As the purpose of this regression was to estimate the effects that being put on probation has on graduation time, the main variable being tested was *probation\_ever*. The rest of the variables were included as to not have omitted variable bias. Without their inclusion probation could potentially pick up the partial effects of the other variables leading to an inaccurate estimation. When this happens the zero conditional mean OLS assumption is violated and the estimations become biased. After testing for statistical significance, the final model is as shown:

$$\begin{aligned} \text{Gradin4} = & \beta_0 + \beta_1 \text{GPA\_year1} + \beta_2 \text{GPA\_year2} + \beta_3 \text{totcredits\_year1} + \beta_4 \text{totcredits\_year2} + \\ & \beta_5 \text{probation\_ever} + \beta_6 \text{suspended\_ever} + \beta_7 \text{gpacutoff} + \beta_8 \text{hs\_grade\_pct} + \beta_9 \text{female} + \beta_{10} \text{english} \\ & + \beta_{11} \text{french} + \beta_{12} \text{asia} + \beta_{13} \text{canada} + u \end{aligned}$$

The coefficient of *probation\_ever* gives the probability that a student graduates in 4 years given that they were placed on probation. Just as with the previous model, the estimations of the predicted probabilities were outside of the [0,1] range, necessitating a probit model. The fitted values before and after the probit model are shown in the graphs:



Additionally, like the last model, robust standard errors were used to take into account the heteroskedasticity of the residuals. The average marginal effect of probation was then computed using margins. The results are shown in table 5:

**Table 5: Gradin4 Margins**

	dy/dx	Std. Error	P >  z	95% Confidence Interval	
probation_ever	-.1237605	.0132555	0.000	-.1497408	-.0977801

Being put on probation has a negative effect on whether or not a student graduates in 4 years. If a student is placed on probation in either of their first 2 years at college, then they are predicted to be 12.4% less likely to graduate on time. This result implies that being placed on probation has a long term negative effect on academic performance. However, it is not sufficient enough evidence to prove that probation affects overall performance/success. Further regressions would need to be done in order to further flesh out this question. These regressions would likely need to include variables such as drop out rates, time in school, and whether or not they graduated at all which are not included in the data set.

### **Short Term Academic Performance**

Seeing as graduation time is not the only variable that is indicative of academic performance, another regression was performed, this time with *GPA\_year2* as the explained variable. This regression was designed with the purpose of uncovering the effects that probation has on short term academic performance. Knowing that 20% of all college students find themselves on academic probation after their first year, the effects that probation has on their



GPA in year 2 was put to the test. Unlike the other regressions thus far, this regression was not a linear probability model. Instead it was a standard multiple linear regression. The initial model:

$$GPA\_year2 = \beta_0 + \beta_1 GPA\_year1 + \beta_2 totcredits\_year1 + \beta_3 totcredits\_year2 + \beta_4 age\_at\_entry + \beta_5 probation\_year1 + \beta_6 suspended\_year1 + \beta_7 suspended\_year2 + \beta_8 campus2 + \beta_9 campus3 + \beta_{10} hsgrade\_pct + \beta_{11} female + \beta_{12} english + \beta_{13} french + \beta_{14} america + \beta_{15} asia + \beta_{16} canada + u$$

This model provides the partial effect that the independent variables have on the dependent variable. The regressor of key importance is *probation\_year1*. This shows the effect that being placed on probation after year 1 has on a student's GPA in year 2. Just as before the regressors were tested at the 5% significance level, creating the final model:

$$GPA\_year2 = \beta_0 + \beta_1 GPA\_year1 + \beta_2 totcredits\_year1 + \beta_3 totcredits\_year2 + \beta_4 probation\_year1 + \beta_5 suspended\_year2 + \beta_6 campus2 + \beta_7 hsgrade\_pct + \beta_8 female + \beta_9 english + \beta_{10} asia + u$$

The results of the regression can be seen in table 6:

**Table 6: GPA\_year2**

	Coefficient	Std. Error	P >  t	95% Confidence Interval	
GPA_year1	.6538452	.0054808	0.000	.6431026	.6645878
totcredits_year1	.0411115	.0063608	0.000	.0286441	.0535789
totcredits_year2	.0500133	.0034633	0.000	.0432251	.0568015
probation_year1	.6329298	.0130674	0.000	.6073173	.6585424
suspended_year2	-1.053705	.0174092	0.000	-1.087827	-1.019582
campus2	.1323658	.0077605	0.000	.117155	.1475765
hsgrade_pct	.0036132	.0001309	0.000	.0033567	.0038697
female	.0676203	.0059649	0.000	.0559289	.0793116
english	.0603298	.0068056	0.000	.0469906	.073669

asia	-.0715037	.0114221	0.000	-.0938913	-.0491161
_cons	.1968426	.0309778	0.000	.1361252	.25756

Intuitively, one may assume that being placed on probation would have a negative effect on GPA. However, as the regression shows, being on probation actually has a positive partial effect on second year GPA. Holding all other variables constant, being on probation in year one leads to a predicted .6329 increase in second year GPA. This is likely due to students needing to raise their GPA past the cutoff point or risk being suspended. The  $R^2$  of the model is 0.5655, meaning that 56.55% of the variation in the model can be explained by the regressors.  $R^2$  is also a measure of goodness of fit, meaning that the addition of variables outside of the dataset could make for a more accurate model.

### Conclusion

The variables that have the largest effect on academic probation are student GPA and the GPA cutoff value set by their university. Additionally, while further tests would be needed to give a definitive answer, it seems that being placed on probation negatively affects student performance in the long run. Students placed on probation are less likely to graduate on time. However, probation seems to have a positive effect in the short run. In the year following academic probation GPA is predicted to increase. Until additional tests are done it is hard to answer the question of whether or not probation keeps students in school or pushes them out. However, it seems that academic probation has a positive short run effect that puts students on the right path towards academic success.

### Works Cited

1. Zansler, Shelley. "Academic Probation in College: Why It Happens and How to Get Back on Track." *AccreditedSchoolsOnline.org*, AccreditedSchoolsOnline.org, 2 June 2020, [www.accreditedschoolsonline.org/resources/academic-probation/](http://www.accreditedschoolsonline.org/resources/academic-probation/).

```

1  clear all
2
3  use "C:\Users\ntess\Downloads\canadiancollege.dta"
4
5  sum
6
7  tab sex, gen(gender) //gender1 gender2
8
9  rename gender1 female
10 rename gender2 male
11
12 tab mtongue, gen(mtonegue) //mtongue1 mtongue2 mtongue3
13
14 rename mtongue1 english
15 rename mtongue2 french
16 rename mtongue3 other_lang
17
18 tab birthplace, gen(birthplace) //birthplace1 birthplace2 birthplace3 birthplace4
19
20 rename birthplace1 america
21 rename birthplace2 asia
22 rename birthplace3 canada
23 rename birthplace4 other_place
24
25 drop suspended_ever
26 gen suspended_ever=0
27 replace suspended_ever=1 if suspended_year1==1 | suspended_year2==1
28
29 gen GPA_total = (GPA_year1+GPA_year2)/2
30
31 gen probation_ever=0
32 replace probation_ever=1 if probation_year1==1 | probation_year2==1
33
34 gen hsgrade_pctXtotcredits_year1=hsgrade_pct*totcredits_year1
35
36 reg probation_ever GPA_total totcredits_year1 totcredits_year2 age_at_entry gpacutoff campus2
   hsgrade_pct female english french america asia canada, robust
37
38 reg probation_ever GPA_total totcredits_year1 totcredits_year2 gpacutoff campus2 hsgrade_pct, robust
39
40 predict probation_everhatlpm
41 hist probation_everhatlpm
42
43 probit probation_ever GPA_total totcredits_year1 totcredits_year2 gpacutoff campus2 hsgrade_pct
44
45 predict probation_everhatprobit
46 hist probation_everhatprobit
47
48 margins, dydx(*)
49
50 reg gradin4 GPA_year1 GPA_year2 totcredits_year1 totcredits_year2 age_at_entry probation_ever
   suspended_ever gpacutoff campus2 hsgrade_pct female english french america asia canada, robust
51
52 reg gradin4 GPA_year1 GPA_year2 totcredits_year1 totcredits_year2 probation_ever suspended_ever
   gpacutoff hsgrade_pct female english french asia canada, robust
53
54 predict gradin4hatlpm
55 hist gradin4hatlpm
56
57 probit gradin4 GPA_year1 GPA_year2 totcredits_year1 totcredits_year2 i.probation_ever i.
   suspended_ever gpacutoff hsgrade_pct i.female i.english i.french i.asia i.canada
58
59 predict gradin4hatprobit

```

```
60 hist gradin4hatprobit
61
62 margins, dydx(probation_ever)
63
64 reg GPA_year2 GPA_year1 totcredits_year1 totcredits_year2 age_at_entry probation_year1
suspended_year1 suspended_year2 campus2 campus3 hsgrade_pct female english french america asia canada
65
66 reg GPA_year2 GPA_year1 totcredits_year1 totcredits_year2 probation_year1 suspended_year2 campus2
hsgrade_pct female english asia
67
68
```