Spring 2018

Effects of Interventions on Students' Term 2 Retention

Li Xian Cha

Follow this and additional works at: http://repository.stcloudstate.edu/scsu_data

Part of the Applied Statistics Commons, Higher Education Commons, Institutional and Historical Commons, and the Statistical Models Commons

Effects of Interventions on Students' Term 2 Retention

Li Xian Cha

Contents

Effects of Interventions on Students' Term 2 Retention

Executive Summary

Early studies show a student's sense of belonging at the university affects their willingness to continue enrolling in school. The New-Entering Freshmen in Fall 2017 were given a survey during the late September 2017. The results of the survey will be used in this study. Some of these students received interventions during summer or/ and Fall. The interventions given in summer 2017 were Social Belonging Intervention and Control Group Intervention. The Control Group Intervention was not designed to have any effect, but the Social Belonging Intervention is. Then, some students were intervened by the Advising Center in Fall 2017. The Social Belonging Intervention and Advising Intervention are the two interventions that this study will focus on to find out if there is any effect of the interventions on students' term 2 retention. The intervention indicator variables and students' Social Belonging Index and Academic Belonging Index retrieved from the survey results are used to predict students' term 2 retention probability. The statistical power is also calculated in this study to show how likely the effect will be detected when there is an effect there to be detected.

Introduction

Early studies show a student's sense of belonging at the university affects their willingness to continue enrolling in school. The New-Entering Freshmen in Fall 2017 were given a survey during the late September 2017. The research is being conducted by Dr. Melissa Hanzsek-Brill. Students' Social Belonging Index and Academic Belonging Index are retrieved from the survey results. Some students were intervened during summer or/ and Fall 2017. The intervention given in the summer is called Social Belonging intervention, designed by Professor Jazwinski and Professor Harris in SCSU. Students were randomly assigned to be in the SBI or EAC group. Students in the SBI group were required to write an essay drawing on personal experience in describing how a new student can gain a feeling of belonging while the students in the EAC group were required to write a more general essay, without drawing on personal experience or reflection. The EAC group is also called Control Group intervention in this study, which is not designed to have any effect. However, not all students participated in summer; there was a group of students who were unwilling or unable to participate. The Advising Center Intervention was done in the Fall, after the school started. Dr. Robinson, the Analytics Projects Consultant, computed the term 2 retention probabilities for each student in the dataset based on a model he ran during September 2017. Part of the students with retention probability lower than or equal to 0.88 were given intervention during the Fall 2017.

The main goal of this research is to find out the effect of the two interventions on students' term 2 retention. A new model predicting the probability of term 2 retention will be produced in this research. The possible predictor variables for the logistic regression model are term 2 retention probabilities by Dr. Robinson, Social Belonging Intervention indicator, Advising Center Intervention indicator, social belonging index, and academic belonging index.

Procedure

This study focuses on the New-Entering Freshmen in Fall 2017. Hence, 101 international students in the original dataset were removed for the new NEF dataset. 1384 non-international students were left in the dataset. Ten students in the new file did not answer all the survey questions. This explains the ten students have social belonging index but not academic belonging index. The belonging indexes are scaled from one to five.

Dr. Robinson, the Analytics Projects Consultant, computed the term 2 retention probabilities for each student in the dataset based on a model he ran during September 2017. Part of the students with retention probability lower than or equal to 0.88 are given intervention by the Advising Center during the Fall 2017. These students are students in group 1, group 2 E, or group 3 E. Group 1 is student with term 2 retention probability below 0.8; Group 2 is between 0.8 and 0.84; Group 3 is between 0.84 and 0.88; Group 4 is student with term 2 retention probability above 0.88. It is important to note that 0.88 is the average proportion for the term 2 retention. Then, a list of students in group 1 and part of group 2 and 3 were sent to the Advising Center. These students are given intervention.

Dummy variable indicating the Advising Center Intervention was created in this study. Dummy variables are variables with only two possible outcomes such as yes and no, but they are coded as 0 and 1. 1 indicates students that are given intervention, which are group 1, group 2 E, and group 3 E; 0 indicates students that are not given intervention, which are group 2 C, group 3 C, and group 4. Then, dummy variable indicating Social Belonging intervention was created. 1 indicating students received the Social Belonging intervention or were in the SBI group. Another column will create dummy variable indicating the Control Group intervention. Students that did not receive or unable to participate in the intervention will have 0's in both columns that indicate the Social Belonging and Control Group interventions.

The statistical power of the two interventions predictor variables are calculated in this study (See Appendix 2 for the formula). The statistical power is the likelihood that a study will detect an effect when there is an effect there to be detected. So, when the power is very close to 1, the test is good at detecting the effect.

The logistic regression model are then produced to predict the probability of the NEF's term 2 retention. A logistic regression model is a regression model where the dependent variable is dichotomous, only two categories/ possible outcomes. The logistic regression model will then

predict the probability of either in the categories, which is the probability of term 2 retention in this study. The logit function is used for the probability calculation (Formula 1).

Formula 1:

$$p = \frac{e^{\beta_0 + \sum_{i=1}^{k} \beta_i X_i}}{1 + e^{\beta_0 + \sum_{i=1}^{k} \beta_i X_i}}$$

P is the probability calculated from the logistic regression model, which is the output in this study. K is the number of predictor variables used in the logistic regression model, β_0 is the intercept of the function, which is retrieved from the logistic regression model, β_i's are also retrieved from the logistic regression model, it is the coefficient of the predictor variables and X's represent the predictor variables used to predict the term 2 retention probability.

60 students' GPA are 0. These students are excluded in later analysis because these students are less likely to return to SCSU at their 2nd term.

Results

Students are categorized according to if they received the social belonging intervention during summer: yes or no; students are also categorized according to if they were intervened by the Advising Center during Fall: yes or no (Figure 1). There are 1304 students in this study. 436 (128 + 308) students were given the social belonging intervention in summer. Besides that, 402 students were intervened by the Advising Center in Fall 2017. 128 students got both interventions (Figure 1).

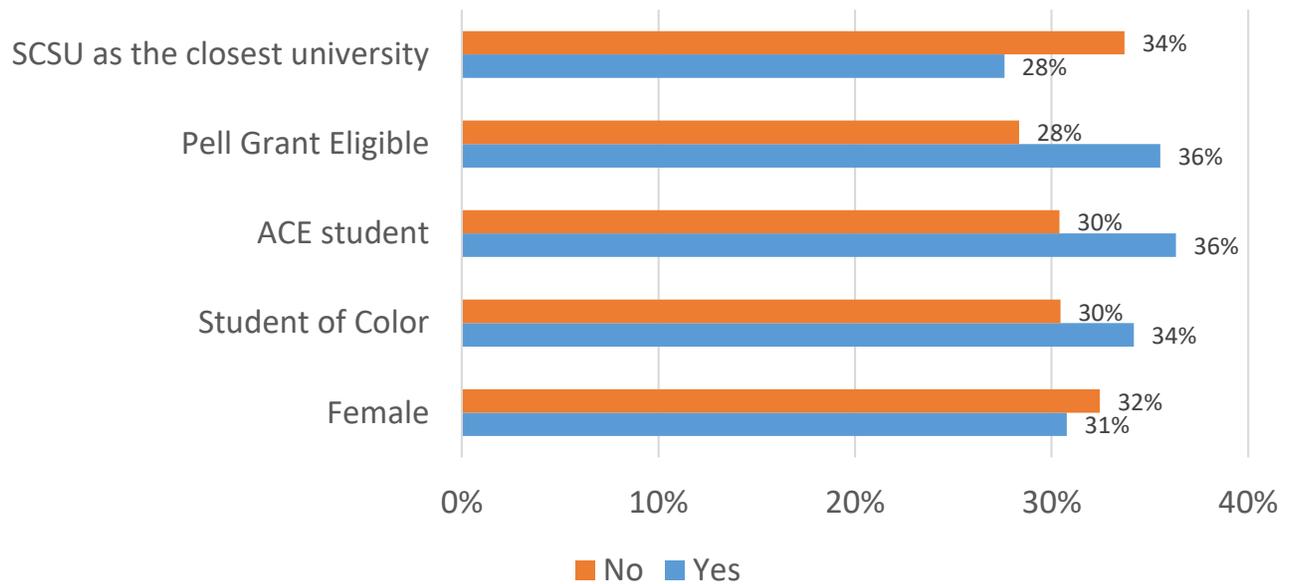*Figure 1: Amount of students that received interventions*
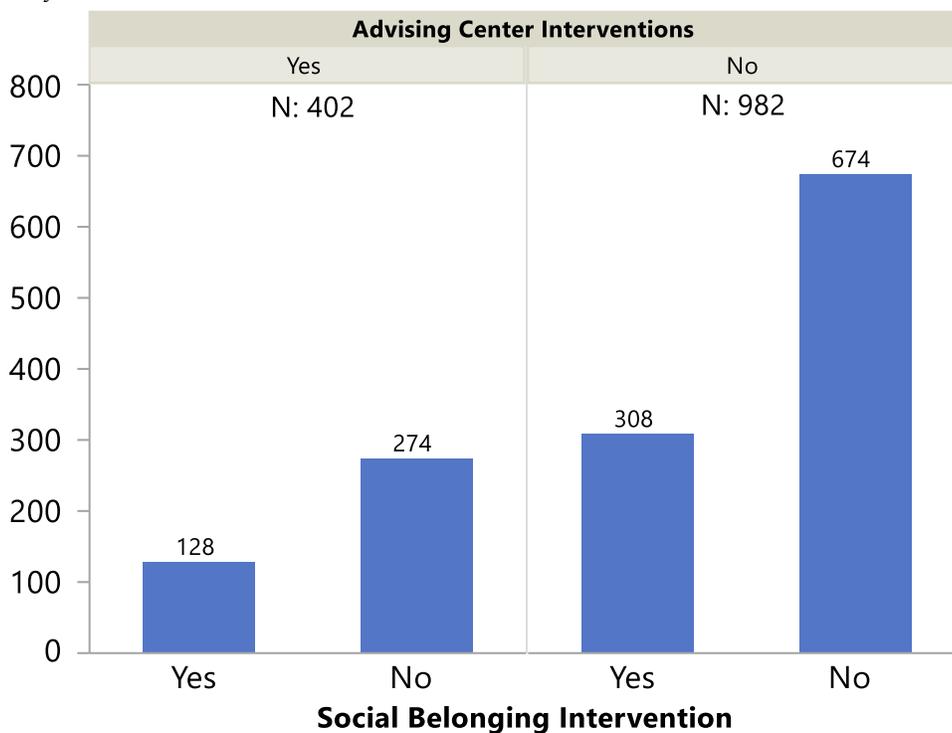
Students are also categorized to other different groups: SCSU as closest university, Pell grant eligible, ACE students, students of color, and female students (Figure 2). Yes or no are the only two outcomes for these categories.

Figure 2: Students' Information vs. Social Belonging Intervention

In the group of students with SCSU is not their closest university to their hometown, 34% of that group of students received the Social Belonging Intervention during summer. 28% of the students in the group where SCSU is the closest university to their hometown were given the Social Belonging Intervention during summer. Overall, 32% of the students were given the intervention in summer. There are differences between the categories, but these differences are not statistically significant. These differences might occur just by chance (Figure 2).

Figure 3: Belonging Levels vs. Social Belonging Intervention
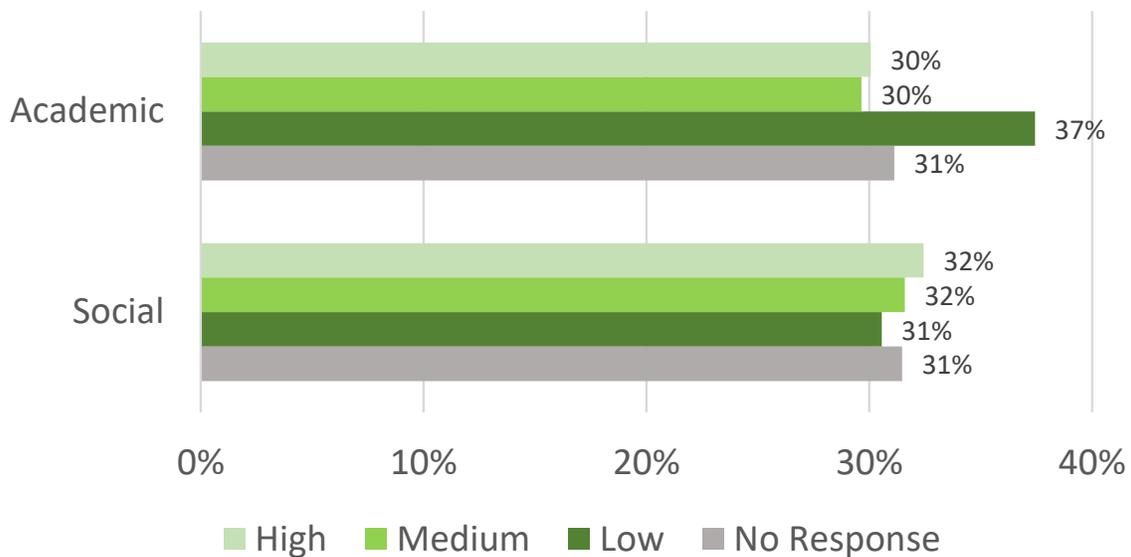
Students are also categorized to different belonging levels. The academic and social belonging levels are shown (Figure 3). The scale for the belonging indexes are from 1 to 5 and there are three belonging levels. 4.5 or above is high, 3.5 to 4.49 is medium, and 3.49 or below is low. Since there are some students who did not response to the survey, they will be in the group of "no response". The students from the low academic belonging level have a slightly higher percentage of students who received the Social Belonging Intervention in summer, which is 37% of them, but this is not statistically significant.

Although the control group intervention is not designed to have any effect on students' term 2 retention, the percentage of students in each categories who received the intervention are shown (Figure 4). Overall, there are 32% of the students received the control group intervention. It is consistent among the categories.

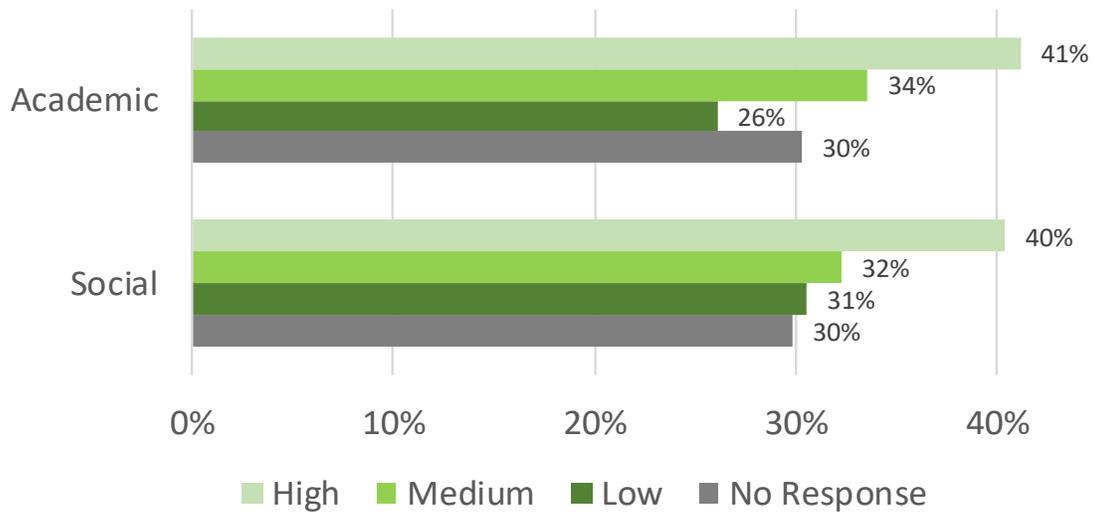*Figure 4: Students Information vs. Control Group Intervention*
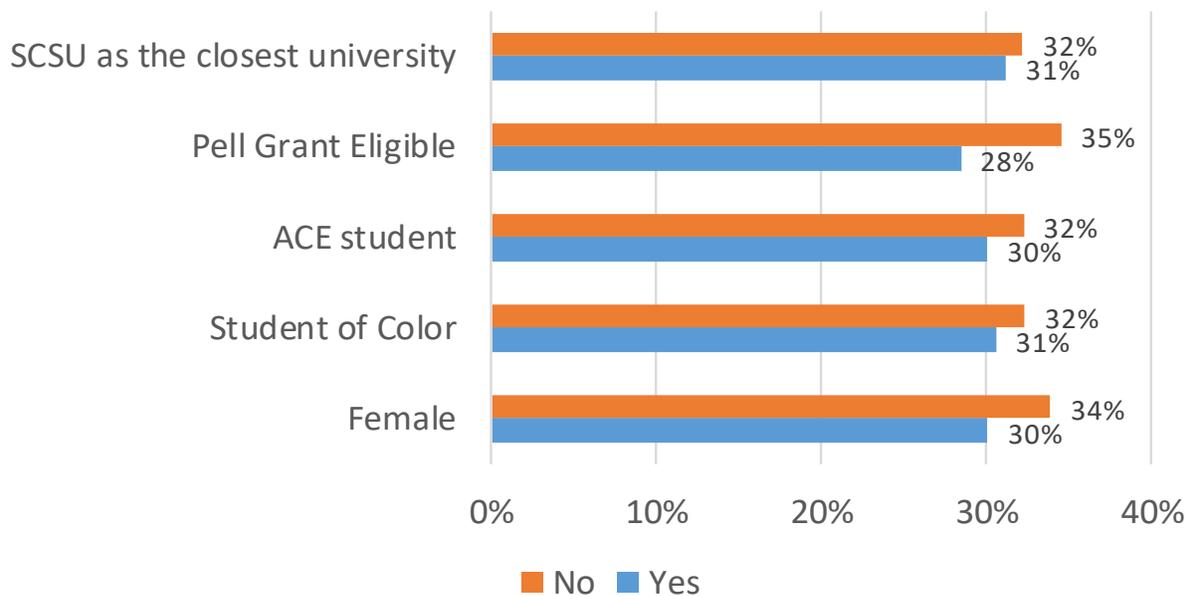
The percentages of students who received the control group intervention are shown in different belonging levels (Figure 5).

*Figure 5: Belonging Levels vs. Control Group Intervention*

The students with higher belonging levels were more likely to receive the control group intervention. Although this intervention is not designed to have any effects, it is possible that there are effects of the intervention since the intervention was done in summer and the survey was given in Fall. However, it might also be random variations since the sample size is small.

The percentages of students that were intervened by the Advising Center during Fall 2017 by different categories were calculated (Figure 6). Since the Advising Center Intervention was given according to Dr. Robinson's list of students from the model he ran to compute term 2 enrollment probabilities, some group of students were more likely to receive the intervention in Fall. The term 2 enrollment probabilities were calculated using some students' information such as ACE student's indicator, Pell grant eligible, housing in SCSU, student of color indicator, SCSU as closest university, student's QPP score and more (Appendix 1). Part of the students with term 2 enrollment probability lower than 0.88 were given intervention during Fall 2017. So, students from certain groups were more likely to receive the Advising Center Intervention since they might have lower term 2 enrollment probability. Overall, 29% of the students were intervened by the Advising Center. ACE students were more likely to receive the intervention because these students are usually in higher risk.

Figure 6: Students Information vs. Advising Center Intervention
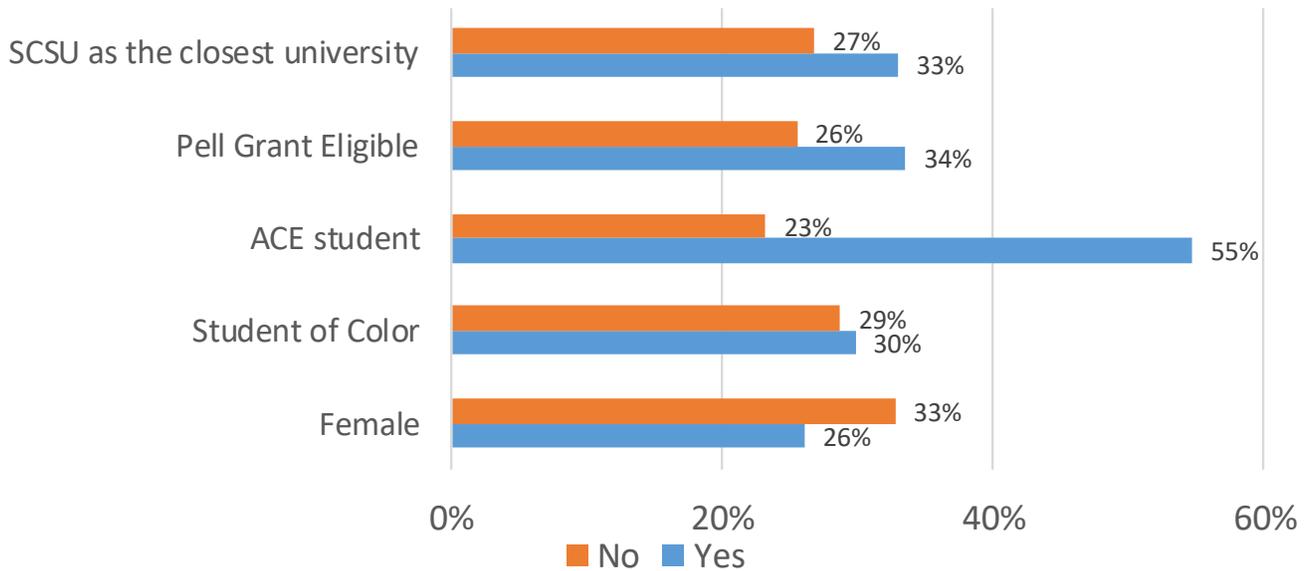
Although the model used to predict the term 2 enrollment probability by Dr. Robinson did not include the belonging indexes, student with low belonging level or did not respond to the survey were more likely to receive the Advising Center Intervention in Fall (Figure 7). There are possible effects of the interventions.

Figure 7: Belonging Levels vs. Advising Center Intervention
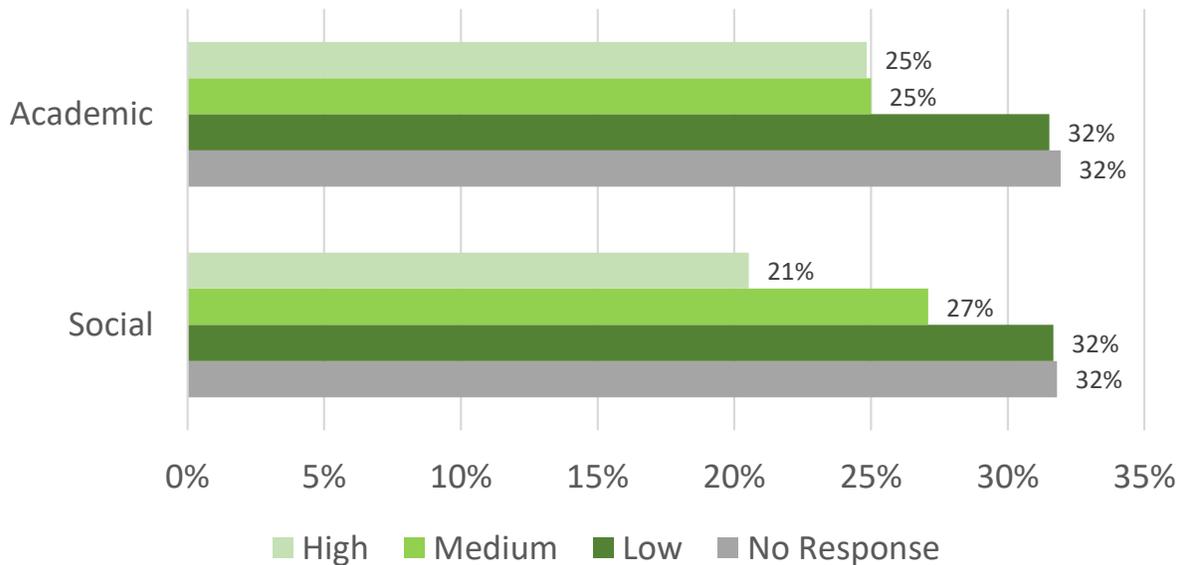
The overall Term 2 retention is 88%. So, the probability calculated from each the predictor variables in the logistic regression model are compared with 88% to show the effect of the predictor variables. The logistic regression model is produced using some students' information and the interventions indicators as predictor variables. There are 13 predictor variables in the logistic regression model to predict term 2 retention probability (Table 1).

Table 1: Logistic regression model to predict NEF's term 2 retention

Term	Estimate	Std. Error	P-Value	Probability	Effect
Intercept	-2.29	2.01	0.26		
Social Belonging Index	1.66	0.29	<.0001	98%	9%
Academic Belonging Index	-0.67	0.35	0.06	79%	-9%
SCSU as the closest university	0.84	0.37	0.03	95%	6%
Student of color	-0.34	0.36	0.34	84%	-4%
ACE students	0.32	0.41	0.43	91%	3%
QPP Score (10 points change)	0.30	0.16	0.06	91%	3%
Advising and registration month	-0.23	0.18	0.20	86%	-3%
Social Belonging Intervention	-0.19	0.36	0.59	86%	-2%
Advising Center Intervention	-0.13	0.36	0.72	87%	-1%
Female students	-0.10	0.33	0.75	87%	-1%
Control Group Intervention	0.05	0.37	0.89	89%	1%
College of Science and Engineering students	0.05	0.39	0.90	89%	0%
Pell Grant Eligible	-0.04	0.32	0.91	88%	0%

Every one unit increases of the each predictor variable will lead to the increase by the "Effect" on the probability of term 2 retention. For example, when student A has a social belonging index as 3 points and student B has social belonging index as 4 points. While assuming all the other variables remain constant, student B will have 9 percentage point more of term 2 retention probability than student A. The one point more of the social belonging index will lead to a 9 percentage point more on the term 2 retention probability. Some variables are dummy variables, where there are only two categories: yes or no. These variables are: SCSU as the closest university, student of color, ACE student, Social Belonging Intervention, Advising Center Intervention, Female students, Control Group Intervention, College of Science and Engineering students, and Pell Grant Eligible. Unlike the other variables, the QPP is 10 points change instead of 1 point because 1 point change will have very small effect, which is less

meaningful. QPP is based on a student's high school GPA, high school class rank percentage, and the composite ACT score. When a student has 10 points higher than another student in QPP while all the other variables remain constant, the student with higher QPP will have 3 more percentage point of the term 2 retention probability.

This study focuses on the interventions, but the interventions as predictor variables in the logistic regression model are showing negative and non-significant effect term 2 retention probability. The effects of the interventions are too small, which is not significant.

A simpler logistic regression model is produced to predict the New-Entering Freshmen's term 2 retention (Table 2). This model included the term 2 enrollment probability computed by Dr. Robinson. This term 2 enrollment probability was computed using some students' information such as ACE student indicator, Pell grant eligible, housing in SCSU, student of color indicator, SCSU as closes university, student's QPP scores and more. The term 2 enrollment probability is from 0 to 1. It is rescaled to be 0.25 point change instead of 1 point change because a 1 point increase from 0 to 1 means student is predicted to enroll in term 2 according to Dr. Robinson's model and it is less comparable. Besides that, to avoid high correlation between the predictor variables, when T2 enrollment probability is used as one of the predictor variables in the logistic regression model, the variables used to compute the probability would not be used. The Social Belonging Index and Academic Belonging Index are highly correlated. This explains why Social Belonging Index is showing positive coefficient/ estimate, but Academic Belonging Index shows negative coefficient (Table 1). The correlation between the two is 0.6508, showing a high positive relationship between the two. So, one must be removed from the logistic regression formula.

Table 2: Simpler logistic regression model to predict NEF's term 2 retention

Term	Estimate	Std. Error	P-Value	Probability	Effect
Intercept	-7.93	2.45	0.00		
T2 Enrollment Probability (0.25 point change)	1.61	0.65	0.01	97%	9%
Social Belonging Index	1.31	0.20	<.0001	97%	8%
Social Belonging Intervention	-0.16	0.33	0.62	86%	-2%
Advising Center Intervention	0.05	0.36	0.90	89%	0%
Control Group Intervention	0.04	0.34	0.90	89%	0%

Both interventions have very small effects on the NEF's term 2 retention probability (Table 2). Moreover, the effect of Social Belonging intervention on the NEF's term 2 retention probability is negative, which is unlikeable. The evidence of success of interventions was not able to detect. The p-values of the three interventions predictor variables are also non-significant.

It is important to find out the effect of the interventions. The statistical power calculated in this study using the information from the logistic regression produced will give a better understanding about the effectiveness. Statistical power is the likelihood that a study will detect an effect when there is an effect there to be detected. When the power is very close to 1, the test is good at detecting an effect. A 3 percentage point change on the term 2 retention probability will be tested for the statistical power.

Table 3: Statistical power of the interventions variables

Interventions	Estimate	Std. Error	P-Value	Power (+3%)	Power (-3%)
Social Belonging	-0.16	0.33	0.62	16%	12%
Control Group	0.04	0.34	0.90	16%	12%
Advising Center	0.05	0.36	0.90	15%	11%

The power is only 16% to detect an increase of 3 percentage point (from 88% to 91%) in the term 2 retention probability when students were given the Social Belonging Intervention in summer 2017. The power is also small to detect a decrease of 3 percentage point (from 88% to 85%) in the term 2 retention probability when students were given the intervention in summer. Same goes to the Control Group Intervention. Besides that, the power to detect 3 percentage point change when students were intervened by the Advising Center during Fall 2017 is only 15% or 11%. The powers are small, meaning statistical significance is not likely for this effect size. A good power would be 80% or above.

Students with GPA = 0 were removed while producing a new logistic regression model to predict New-Entering Freshmen's term 2 retention It is a simple logistic regression like Table 2 to predict the New-Entering Freshmen's term 2 retention but without students with GPA = 0. These students are students with GPA higher than 0 (Table 4).

Table 4: Simple Logistic regression model to predict NEF with GPA>0's term 2 retention

Term	Estimate	Std. Error	P-Value	Probability	Effect
Intercept	-6.17	2.83	0.03		
Social Belonging Index	1.37	0.21	<.0001	98%	6%
T2 Enrollment Probability (0.25 point change)	1.08	0.76	0.16	97%	5%
Social Belonging Intervention	-0.18	0.36	0.61	90%	-1%
Advising Center Intervention	0.11	0.39	0.78	93%	1%
Control Group Intervention	-0.07	0.36	0.85	91%	-1%

Students with GPA = 0 were dropped out of school or suspended in the end of the semester. These students are less likely to return to SCSU. Since these students were removed, the effects of the interventions might be detected better through the logistic regression model to predict NEF's term 2 retention probability. However, the two interventions are still showing almost no effect on the NEF's retention (Table 4). The effects are relatively small and less reliable since the p-values are large.

The power to detect 3 percentage point change on the term 2 retention probability while interventions were given to the students were calculated (Table 5). However, the power are still small to detect 3 percentage point change in this study even the students with GPA = 0 were removed.

Table 5: Statistical power of the interventions variables (GPA>0)

Interventions	Estimate	Std. Error	P-Value	Power (+3%)	Power (-3%)
Social Belonging	-0.18	0.36	0.61	28%	16%
Control Group	-0.07	0.36	0.85	28%	16%
Advising Center	0.11	0.39	0.78	24%	14%

Conclusion

Comparing the percentage of students under the Control Group Intervention with their belonging levels, the intervention may affect students' retention to some extent. However, the Control Group Intervention was not designed to have any effects. When the Control Group Intervention indicator is used as a predictor variable in the logistic regression model, it does not show a significant impact on the NEF's term 2 retention either. The power is too small to detect 3 percentage point change on students' term 2 retention probability while students received the Control Group Intervention. A low statistical power has a declined chance of detecting the effect. The small sample size in this study might result in the weak power. The Social Belonging Intervention and Advising Center Intervention were designed to improve students' term 2 retention but these interventions are not showing statistical significance. While the interventions were given to the students, the power to detect 3 percentage point change on the term 2 retention probability are small. The effects of the interventions are not able to be detected even if the effects exist. Although the interventions might or might not have an effect on students' term 2 retention, this study is unable to prove it due to small sample size and weak power. However, the power can be improved if the sample size of the study is 8 times larger while the effect sizes of the interventions remain constant (Appendix 4). The interventions were poorly carried out, and the students who participated were not very engaged and not taking in the information. A better intervention should be designed next time to improve the effectiveness.

Appendix

Appendix 1: T2 Enrollment Probability Formula

$$Link = (-0.65338737851278) + 0.00118374435070832 * AdmitDaysBeforeTerm$$
$$+ 0.261988157802486 * AR\_Missing\_Flag + 0.0266555613408153 * QPP2$$
$$+ 0.109765031249571 * ACEFlag + 0.552760285290723 * IntlFlag$$
$$+ 0.234732132095622 * SOCFlag + 0.0843419726559161 * ClosestToSCSU$$
$$+ -0.0990442946603616 * PellEligFlag + 0.106752949157006$$
$$* FirstGenFlag + 0.243285814598277 * HousingFlag + 0.136642245212192$$
$$* YRTRCreditsGT14 + -0.82116432336717 * YRTRCreditsLT12$$
$$+ -0.121576775480201 * AR\_August\_Flag$$

$$T2\ Enrollment\ Probability = exp(Link)/(1 + exp(Link))$$

Appendix 2: Power Calculation

	A	B	C	D	E	F	G
1			Logit	Difference from Null	Std. Error	Z-Score	Power
2	**Higher Alternative**	=B3+0.03	=LN(B2/(1-B2))	=C2-C3	SE	=ABS(D2/E2)	=NORM.S.DIST(F2-1.96,TRUE)
3	**Null**	p	=LN(B3/(1-B3))				
4	**Lower Alternative**	=B3-0.03	=LN(B4/(1-B4))	=C3-C4	SE	=ABS(D4/E4)	=NORM.S.DIST(F4-1.96,TRUE)

Appendix 3: Logistic regression model to predict term 2 retention probability (GPA>0)

Term	Estimate	Std. Error	P-Value	Probability	Effect
Intercept	-2.22	2.19	0.31		
Social Belonging Index	1.68	0.30	<.0001	98%	7%
Academic Belonging Index	-0.60	0.37	0.11	86%	-6%
SCSU as the closest university	0.95	0.41	0.02	97%	5%
Student of color	-0.27	0.39	0.48	90%	-2%
Social Belonging Intervention	-0.24	0.38	0.54	90%	-2%
QPP Score (10 points change)	0.15	0.18	0.40	93%	1%
Female students	-0.12	0.35	0.74	91%	-1%
Advising Center Intervention	-0.11	0.39	0.78	91%	-1%
Control Group Intervention	-0.09	0.39	0.82	91%	-1%
Pell Grant Eligible	0.09	0.35	0.79	93%	1%
College of Science and Engineering students	-0.08	0.41	0.85	91%	-1%
Advising and registration month	-0.07	0.21	0.72	91%	-1%
ACE students	0.00	0.46	0.99	92%	0%

Appendix 4: Recalculated power when sample size is 8 times larger

The power is recalculated assuming the sample size is 8 times larger while the effects of the predictor variables remain the same. The power is a lot better when the sample size is 8 times larger.

For all students in this study:

Interventions	Estimate	Std. Error	P-Value	Power (+3%)	Power (-3%)
Social Belonging	-0.16	0.33	0.62	80%	61%
Control Group	0.04	0.34	0.90	78%	59%
Advising Center	0.05	0.36	0.90	73%	53%

Students with GPA > 0:

Interventions	Estimate	Std. Error	P-Value	Power (+3%)	Power (-3%)
Social Belonging	-0.18	0.36	0.61	97%	78%
Control Group	-0.07	0.36	0.85	97%	71%
Advising Center	0.11	0.39	0.78	94%	78%

Appendix 5: Data column meaning

Column Name	Column meanings
TechId	Student's ID
YRTR	Year Term
Gender	Student's gender
GenderFlag	Female = 1, Male = 0
SOC	Student of Color
SOC_Flag	SOC = 1, non-SOC = 0
Intl	International Student
Intl_Flag	International Student = 1, non-International Student = 0
ARMonth	Advising and registration month
QPP2	QPP score
ACE	ACE students
Pell_Elig	Pell Grant Eligibility
Pell_Flag	Pell Grant Eligibile = 1, Pell Grant not eligibile = 0
ClosestToSCSU	SCSU as the closest university
ClosestFlag	SCSU is the closest university = 1, SCSU is not the closest university = 0
T2 Enr Prob	T2 Enrollment Probability
EAC-Ctrl, SBI-Trt	Social Belonging Intervention Group
Retention Gp	Retention Group
T2_Enr Day 1	Enrolled during Day 1
T2_Enr Day 9	Enrolled during Day 9 of class
Q Order	Question order
COSE	College of Science and Engineering students
COSE_Flag	College of Science and Engineering students = 1, not College of Science and Engineering students = 0
Social Belonging Index	Social Belonging Index
Social Belonging Group	Social Belonging Group
Academic Belonging Index	Academic Belonging Index
Academic Belonging Group 1	Academic Belonging Group 1
Academic Belonging Group 2	Academic Belonging Group 2
Academic Belonging Group 3	Academic Belonging Group 3 (Final) - this is used
Cumulative_Att_Cr	Cumulative attempt credit hours
Fall17_Att_Cr	Fall 2017 attempt credit hours
Cumulative_Earned_Cr	Cumulative earned credit hours
Fall_Earned_Cr	Fall 2017 earned credit hours
Cumulative_GPA	Cumulative GPA

Fall17_GPA	Fall 2017 GPA
GPA=0?	Is GPA = 0

Appendix 6: Statistics theory about power and others

1. The two types of errors in hypothesis tests

There are two hypotheses in the hypothesis test. They are the null hypothesis and alternative hypothesis. The null hypothesis is the default position saying that there is no relationship between two things measured in the test, or there is no difference between the groups; the alternative hypothesis is what you might believe to be true of hope to prove true in most cases, the negation of the null hypothesis. The definition of type 1 error is rejecting the null hypothesis when the null hypothesis is true. It is also known as the false positive, a result that indicates a given effect or condition is present when it is not. A threshold, denoted as α, is pre-specified to limit the probability of a type 1 error. It is called the level of significance, the largest value of α accepts. The definition of type 2 error is failing to reject the null hypothesis when the alternative hypothesis is true. It is also known as the false negative in a test checking for a single condition with a definite result of true or false. When a type 1 error is more serious than a type 2 error, choose α to be smaller (usually $\alpha=0.05$), say 0.01; When a type 1 error is less serious, choose α to be larger, say 0.1.

2. The meaning of p-values and how to best use them

P-value is used all over statistics, from t-tests to regression analysis. P-values are often being used to determine statistical significance in a hypothesis test. When a very small p-value is obtained in the test, say less than the level of significance, α, the null hypothesis will be rejected, concluding a result statistically significant. This is, high p-value means your data are likely with a true null and low p-value means your data are unlikely with a true null. It is not the error rate or the probability of making a mistake by rejecting a true

null hypothesis (type 1 error). Lack of statistical significance does not indicate that the effect size is small. One must look at the confidence interval to determine whether it includes effect sizes of importance.

3. The meaning of power and how to access it

To calculate power: 1 minus the probability of type 2 error. It is the likelihood that a study will detect an effect when there is an effect there to be detected. The power of a test is the probability that the right decision is made when the null is not correct. So, when the power is very close to 1, the hypothesis test is very good at detecting a false null hypothesis. The significance level of the test, sample size, and variability in the measured response variable can affect the power calculation.