Retention Rates at West Chester University: A Study on Retaining First Year, First Semester Students at West Chester University

Sydney Treiman, West Chester University

I.   Introduction

Student retention at any University or Community College is dependent on multiple and varying factors and place a large role in the success of any university. Retention rates allow universities to evaluate if their programs, professors, and overall campus life is successful. This also factors into the ability for new students, specifically first semester, first year students to acclimate to the new environment and living situation. Information will be provided on the variables that contribute to student retention and the findings that go along with each variable. From previous findings, those who had a lower retention rate included individuals who had a sense of non-belonging, higher stress levels than usual which was due to the new environment and work load. This can attribute to anything from being in an unfamiliar setting to not having your family around like they used to be. Many students leave due to personal reasons such as failing classes, getting in trouble with the law or just being home sick. It is difficult for students to leave home after 18 years and move away from their comfort zone. This is a difficult transition for a student who's never done anything like this before. It is evident that the key to being considered a retained student includes making friends/having a sense of belonging, joining clubs and organizations, or just finding a hobby that others enjoy, too. West Chester University offers various clubs, organizations, and student involvement for new students. Many organizations promote themselves to first year, first semester students to get them involved from the start of their college career.

Some variables have less of an impact on retention rates than what one might think otherwise. The question that will be answered is as follows; what variables have the most impact on retention rates at West Chester University? These variables play an important role in determining if a student will return the next semester.

Retention in this study is defined as a first year, first semester student who stays at the university consecutively for three semesters. In this case, the student would need to be at West Chester University for fall 2015, spring 2016 and fall 2016. This would be one full year at the university and this would be a student who's considered to be retained. There will be evaluation of West Chester University's ability to retain students using variables such as the number of day's students are on campus and their living situation. West Chester University does, in fact, have a generally high retention rate which is promising for future students who attend.

II.   Literature Review

A main resource that has relevance on this topic was written by Vincent Tinto in 2006. It is titled *Journal of College Student Retention: Research, Theory & Practice.* Volume 8. *Research and Practice of Student Retention: What Next?* In this Volume, the author investigates reasons to support the information regarding retention rates. The author begins with a comparison between student retention from then and now, which was about 40 years ago. He explains how those who did not graduate were viewed as less capable, less motivated and less willing to utilize what

colleges had to offer. Students were questioned to determine why they did not return. However, Tinto explains and examines that it is the environment and institution in which the student attends that contributes to retention rates.

There is an examination of the data that used to be collected for student retention data and what it suggested. Most of the early work came from quantitative studies of mostly residential universities where the majority of students lived on campus and students of majority backgrounds.

As Tinto begins, he mentions how crucial student involvement is in the first year which has a large effect on if the student returns the following year. This is related to the research conducted as the primary focus is determining what causes a student to be retained compared to those who are not. The importance of freshman events such as extracurricular activities, freshman seminars, and living situation play a major role in first year retention. There have been multiple changes to the study of retention rates. Some of these changes include looking at culture, race, socioeconomic status and involvement on campus. At one point, this type of information was not considered for student retention. In addition, Tinto explains that 2 or 4-year colleges are now being accounted for, and, whether students live on or off campus. This meaning that there is a comparison between the two-year college and the four-year college. Also, that being residential (on campus) and being non-residential (off campus) has an effect. Finally, it is determined that faculty are a key to increasing student retention. With faculty that's engaged and concerned for their students, there would be better results. Besides these points, Tinto addresses other aspects of student retention.

The author points out three lessons that are crucial to student retention; institutional action, program implementation, and student income.

The first, institutional action, has to do with knowing and understanding why students leave. This becomes a job for the institutions to find out what they can do to help students stay and succeed at their university. A major component being social, and academic integration. While this helps those researching student retention, it doesn't help institutions to determine how to have first year students both academically and socially integrated. While there might be programs for this, a student can't be forced to participate. There is information regarding high school and family experiences which contribute to retention rates, too. Knowing this type of information can help institutions to have programs for students with similar backgrounds. This gives them a place to feel at home, without actually being home. It falls on the institutions to attempt to create this, but, the institutions don't know how to. "What is needed and what is not yet available is a model of institutional action that provides guidelines for the development of effective policies and programs that institutions can reasonably employ to enhance the persistence of all their students. Such a model would have to describe in a coherent, interrelated manner, how different institutional actions impact student persistence in varying ways and degrees in both two- and four-year settings" (pg. 6).

The second point of interest is program implementation. While it is possible to identify effective action, it is more of a challenge to put it to use. Beginning a program is easy, while seeing it endure is hard. Part of the reason for low student retention has to do with the fact that some programs might be implemented but don't endure. Thus, leaving the student(s) with no outlet or source of involvement. Unfortunately, the only involvement tested in the data is that of the

number of visits to the student recreation center. If a student has no involvement, it is unlikely the individual is meeting others. This is crucial to the development of a first year student.

Student income plays another major role in student retention. Those students coming from low socioeconomic status's are less likely to attend a 4-year institution compared to a high socioeconomic student's status. Tinto explains how there is less socioeconomic diversity than racial, but there is economic diversity at the most selective colleges and universities. Part of the data gathered from West Chester includes both student and parent accrued gross income.

Tinto should've included some more statistics and graphs to demonstrate the findings. This is something that would have benefited the reader. Visuals help to show a representation of the points being brought up for discussion.

Another main source of data is written by Patrick O'Keefe in 2013 titled *A Sense of Belonging: Improving Student Retention.* Volume 47, Issue 4. *College Student Journal.* O'Keefe gives information about why student retention has become a major problem for higher education world-wide. Part of the reason for students to leave college have to do with feeling as if they don't belong, have a feeling of rejection, and may not have the ability to adjust to normal academic challenges that go along with college life. There are certain groups that are listed who are considered at risk of non-completion of college; ethnic minorities, academically disadvantaged, students with disabilities, those of low socioeconomic status, and probationary students. Those who are also a first generation college student pose a risk of not completing four years, too. Part time students have an even lower retention rate compared to full time students as they aren't on campus as often, and aren't as involved. In the data obtained from West Chester University, information will be displayed regarding ethnicities but not the other factors that are mentioned.

Mental illness plays a role in retention rates as those with mental illness are at a higher risk of leaving the institutions. Because of the increase in mental illness at universities, it becomes a multi-faceted issue for higher education institutions as it is occurring more and more often. Some institutions don't have the appropriate resources to help those with mental illness which leads to an even bigger increase in students exiting the university.

Student retention has effects on the institution that can cause major issues. There can be a lack in revenue, lack of funding, and others. O'Keefe gives strong evidence for the major points mentioned in this journal. He gives explanation as to why students are leaving and solutions for what might encourage them to stay. If there were graphs with statistics about those who are more at risk, it would give a better argument for his statements.

The approach for this research varies from these authors as raw data is used from West Chester University and am not able to gather data on non-quantitative aspects. This would include the information on students not feeling a sense of belonging or if they have disabilities. Other sources would be needed to gather information regarding a sense of belonging or disabilities. However, I'm able to deliver information on if they're part time students, and the ethnicities which plays a large factor in the above findings for student retention.

III.   Methodology

The data involved in this research paper includes one full year from fall 2015, spring 2016 and fall 2016 from West Chester University. This data is cross sectional and is a measurement of student retention. There are 2,917 observations included in this data set. The steps to condense the data involved stacking three sets of information from the three separate semesters. Each file had to be manipulated and then condensed to ensure there was only one record of the student for the final data set.

The limitations to this data is that there was no way to survey these students as they're anonymously given to me. If it was possible to survey these students some possible questions that could be included is; did you feel a sense of belonging freshman year? Were your stress levels higher than usual? How many clubs/organizations did you join freshman year? These questions can help with the data to determine if the information from the other researchers connect to the student's feelings at West Chester University. There are eight major variables used in this research. The first four variables are dummy variables. The first variable is "female". In the data, if the student is a female they're given the value of 1 and if they're male, 0. The next variable is named "living". The data had information regarding if a student lived on campus, or was off campus. If a student is off campus, they're likely a commuter and not just walking a few blocks to the university. This is rare for a freshman. In this case, if a student lived on campus (university student housing or traditional) they were given the value of 1, and if they did not live on campus they're given the value of 0. The third variable is named "home". If the student lives in Pennsylvania (the location of West Chester), they were given the value of 1 and if they live anywhere besides Pennsylvania, they're given the value of 0. The next variable is "race". West Chester University is a predominantly white institution, so, in this case students who are white were given the value of 1, and those who aren't, 0. The next variable is an important one as it is the number of days a student spends on campus ("daysoncampus"). It can be assumed that this number is the number of days a student has class as there's no way to track if a student is actually in class or not. "Parent_AGI" (parent accrued gross income) is the next variable. This also has a value given as this is what's reported when students apply to attend this university. The age of the student is another variable that's crucial in this research. The "age" is restricted in this study from the ages of 18-21. The final variable used in this study is "countsum". This is representative of the number of times in total a student went to the student recreation center over the course of three semesters studied at West Chester University.

Some of these variables have a larger impact than one would expect. However, after gathering the data and looking at the outcome, the regressions make sense for the situation of a first year, first semester freshman.

The descriptive statistics can be referenced in the appendix in table 1. There are six variables included, some of which aren't used in the regressions. However, those variables that are dummy variables can't have descriptive statistics as they only have two values, being 0 and 1. Days on campus which is serves as a very important variable is used and this has a range between zero and five. This being a student can have a maximum of classes up to five days a week.

The information about what gives students more likely of a chance of being retained will be discussed below.

IV.   Results

The coefficients that are not statistically significant will not be discussed in this research paper. This is because they ultimately have no effect on if a student is retained at this institution. These variables include; being a female, the students race and the parent's income. These three variables have no effect on student retention. Both the female and race coefficient is a dummy variable. All three of these variables have p-values of above 0.00 making them not statistically significant.

The model used for regressions in this study is the linear probability model. This is simply where an OLS (ordinary least squares) is run and the dependent variable, which is retained, is a dummy. The students who attended West Chester University consecutively for three semesters (fall 2015, spring 2016, fall 2016) were given the value of 1, and those who did not were given the value of 0. The equation for this is as follows;

$$D_i = \beta_0 + \beta_1 X_{1i} + \varepsilon_i$$

Here it is shown that D_i is a dichotomous variable of value 1 if a student is retained, and 0 if they are not. X_i is a vector of independent repressors including female, living, home, race, DaysOnCampus, Parent_AGI, age and countsum. Since the model employs cross section data, the r-squared and adjusted r-squared are often relatively low, but many coefficients are significant nevertheless.

After much research, multiple ideas can be concluded with how the variables affect retention at West Chester University. Referencing the appendix, table 2, it can be observed that living on campus increases the likelihood of being retained by 17%. This makes sense as those who live on campus will see their peers every day of the week, including weekends. These are the individuals who live in a dorm and will interact with students on a regular basis. These students have the ability to then build relationships, which in turn will encourage them to return the next semester, and the one after that. For the" home" variable, those who are residents of Pennsylvania have a 7.8% chance of being retained compared to an individual who is not. This could be due to cost of tuition for being out of state or that it is a farther distance from home. Being a farther distance from family can cause students to leave school, thus, not being retained. A student who spends an additional day on campus has a 5.9% higher chance of being retained compared to a student who isn't on campus. This ties into living in a dorm, too. For a student who lives in the dorms, and has classes every day, they're constantly socializing and becoming familiar with their peers. Unfortunately, for a student who commutes and is only on campus twice a week, they're less likely to be retained. But, this does make sense. If students are spending most of their time on campus, they would be retained compared to someone who doesn't. Age plays a huge role in the statistics on if a student is retained or not. For every year older the student is, they're 25% less likely to be retained. This, again, makes sense. Most first semester freshman are either 18 or 19 years old. When a student comes in and is 20 or 21 years old, they might not be able to relate to the other freshman who are only one or two years younger. Unfortunately, they have a less likely chance of being retained compared to the typical aged freshman. The final variable used in this OLS regression is "countsum". Students who attend the student recreation center are 0.1% more likely to be retained compared to students who don't attend. The other variables included in this regression in table 2 are not statistically significant. Parent_AGI is representative of the parents accrued gross income. In this instance, it

has next to no effect on is a student is retained. It is relatively normal for students to take out student loans so having their parents pay for college isn't necessary.

Referencing table 3 in the appendix, the difference between this and the regression in table 2 is that there is a squared variable. The squared variable here is "countsum". This variable is representative of the number of times a student attended West Chester University's student recreation center. In table 2, the student was 0.1% more likely to be retained if they attended the student recreation center compared to a student who did not attend. However, when the variable is squared, it is showing that attending the recreation center has no effect on if a student is retained or not. This was surprising as it is usually assumed that students who exercise are more likely to be retained and have better grades. In this case, the data is showing something different. If a student goes double the amount of times they actually went, it has no effect on if they return the following semester.

Table 4 in the appendix is representative of a logit model. This is used as an estimation technique for equations where the dependent variable (retained) is a dummy variable. The equation used for the logit model varies from that of an OLS model. The equation used for the regression is as follows;

$$D_i = 1 \div 1 + e^{-[\beta_0 + \beta_1 X_{1i} + \beta_2 X_2 + \varepsilon_i]}$$

The interpretation of the coefficients for the logit equation is much different than the OLS regression because the equation is very different. The statistically significant coefficients in this regression are living, home, daysoncampus, age, and countsum. Students who live on campus are 83.1% more likely to be retained than students who commute. The students who live in Pennsylvania are 49.9% more likely to be retained compared to students who live outside of this state. A student who spends an extra day on campus compared to a student who doesn't is 32.2% more likely to be retained. Every year older a student gets, they're drastically less likely to be retained. Lastly, a student who attends the student recreation center is 0.9% more likely to be retained compared to a student who doesn't. The findings in the logit model are much different than that of the OLS. The percentages increased in this model.

The last regression contains a variable that is a log. A log is an inverse of an exponential function. In this regression, the function that had the natural log taken is "cumgpa" which is the cumulative GPA for the students involved in this study. The statistically significant coefficients in this regression include the same as both the logit model and the OLS model. For students who live on campus, they're 14.1% more likely to be retained compared to students who are off campus for their first semester. If a student lives in Pennsylvania, they're 6.4% more likely to be retained compared to students who live somewhere other than Pennsylvania. For every additional day a student spends on campus, they're 3.7% more likely to be retained compared to students who aren't on campus often. Every year older a student gets, they're 24.1% less likely to be retained. A student who attends the recreation center is 6% more likely to be retained compared to someone who doesn't go. Lastly, the log variable, for every point higher a student's GPA is, they're 26.9% more likely to be retained.

After multiple regressions and gathered data, it can be concluded that 72% of first year, first semester freshman are retained from this data set. This is specific to the 2,917 anonymous students included in the data. The coefficients used in this study are logical and make sense.

V.   Conclusion

The main findings in this study are that a student who is on campus often is much more likely to be retained. This directly correlates to living on campus, having classes every day, and having the student's home state be Pennsylvania. This is important because with this information, universities can find a way to ensure students who aren't living on campus or aren't at school often have a sense of belonging to ensure they return the next semester. While some universities have organizations such as "off-campus commuter association", those who don't live on campus might not feel as connected as those who do. It was surprising that race, gender and parent income had next to no effect on student retention. It can be concluded that in this data there is no serial correlation. Universities need to continue to offer options for both students living on and off campus to increase retention rates at any university.

VI.   References

Tinto, Vincent, ed. 2006. *Journal of College Student Retention: Research, Theory & Practice.* Volume 8. *Research and Practice of Student Retention: What Next?* Syracuse University Pell Institute for the Study of Opportunity in Higher Education. http://journals.sagepub.com/doi/abs/10.2190/4ynu-4tmb-22dj-an4w

National Center for Education Statistics, 2017. *Undergraduate Retention and Graduation Rates. https://nces.ed.gov/programs/coe/indicator_ctr.asp*

Pike, R. Gary and Graunke,S. Steve, 2014. *Examining the Effects of Institutional and Cohort Characteristics on Retention Rates.* https://link.springer.com/content/pdf/10.1007%2Fs11162-014-9360-9.pdf

O'Keefe, Patrick, 2013. *A Sense of Belonging: Improving Student Retention.* Volume 47, Issue 4. *College Student Journal.* *http://web.a.ebscohost.com/ehost/detail/detail?vid=0&sid=79121e9b-fb2f-4487-929f-28bbc4af0323%40sessionmgr4007&bdata=JnNpdGU9ZWhvc3QtbGl2ZSSzzY29wZT1zaXRl#AN=93813989&db=s3h*

VII.   Appendix

Table 1: Descriptive Statistics

Variable	Mean	Standard Deviation	Minimum	Maximum
Student_AGI	$2,348.75	$7,507.13	-$1,547.00	$200,022.00
Parent_AGI	$94,017.16	$86,408.41	$-499,099.00	$1,249,975.00
Age	20.17	0.46	19	22
DaysonCampus	4.59	0.95	0	5
TotalTaken(Credits)	39.78	8.17	2.00	60
TotalPassed (credits)	37.85	9.08	2.00	60

Table 2: Least squares regression on dependent variable "retained"

	Coefficient	Std. Error	Prob.	Statistical Significance
Female	0.004	0.016	0.76	
Living	0.170	0.023	0.00	***
Home	0.078	0.020	0.00	***
Race	0.016	0.020	0.43	
DaysOnCampus	0.059	0.005	0.00	***
Parent_AGI	-0.000	0.000	0.15	
Age	-0.251	0.012	0.00	***
Countsum	0.001	0.000	0.00	***

R-Squared: 0.23
Adj. R-Squared: 0.24

Table 3: Least squares regression on dependent variable "retained", including a squared variable

	Coefficient	Std. Error	Prob.	Statistical Significance
Female	0.00	0.016	0.76	
Living	0.170	0.023	0.00	***
Home	0.078	0.020	0.00	***
Race	0.016	0.020	0.42	
DaysOnCampus	0.059	0.005	0.00	***
Parent_AGI	-0.000	0.000	0.15	
Age	-0.251	0.012	0.00	***
Countsum	0.001	0.000	0.04	**
Countsum^2	0.000	0.000	0.90	

R-Squared: 0.24
Adj. R-Squared: 0.23

Table 4: Logit regression on dependent variable "retained"

	Coefficient	Std. Error	Prob.	Statistical Significance
Female	0.028	0.103	0.78	
Living	0.843	0.138	0.00	***
Home	0.498	0.126	0.00	***
Race	0.075	0.132	0.57	
DaysOnCampus	0.323	0.034	0.00	***
Parent_AGI	-0.000	0.000	0.14	
Age	-1.377	0.081	0.00	***
Countsum	0.007	0.005	0.12	
Countsum^2	0.000	0.000	0.66	

Table 5: Least squares regression on dependent variable "retained", with a LOG variable

	Coefficient	Std. Error	Prob.	Statistical Significance
Female	-0.027	0.016	0.68	
Living	0.141	0.023	0.00	***
Home	0.064	0.020	0.00	***
Race	-0.021	0.020	0.28	
DaysOnCampus	0.037	0.006	0.00	***
Parent_AGI	-0.000	0.000	0.08	*
Age	-0.241	0.012	0.00	***
Countsum	0.000	0.000	0.00	***
LOG(cumgpa)	0.269	0.022	0.00	***

Table 6: Retained

	Retained
Mean	**0.719**
Median	1.000
Maximum	1.000
Minimum	0.000
Standard Deviation	0.449
Skewness	-0.976

KEY

*** is representative of being statistically significant at the 1% level
** is representative of being statistically significant at the 5% level
* is presentative of being statistically significant at the 10% level