Implementation of Course-Based Learning Communities and Living Learning Communities along with the Development of a Simple Python Program for Measuring Retention

Dr. Mary E. Goodwin, University of South Florida

Dr. Goodwin, who has engineering degrees in industrial and environmental engineering, is the Director of Student Services in the College of Engineering at the University of South Florida. She worked in industry for 9 years and over 20 years in higher education focusing on engineering education.

Mr. John Pharo Morgan III, University of South Florida

John earned his M.Ed. in 2003 and has worked as an engineering academic advisor for the past ten years. He is currently finishing his dissertation in Curriculum and Instruction.

Yan Wang, University of South Florida

She is a doctoral student in the Educational Measurement and Research program at USF. Her research interests include the application of statistical methods to the analysis of educational data, test/instrument development, and psychometrics. She received her M.Ed. in Curriculum and Instruction at Boston College.

Mrs. Michelle King, University of South Florida College of Engineering

Michelle King has 15 years working in higher education. She is a user applications specialist, specializing in university student information systems and has been dubbed the College of Engineering’s “data guru” by her colleagues. She is currently finishing her degree in Information Technology.

Mr. Blake A Burton, University of South Florida

Computer Science major graduating in Fall 2015 with a bachelor from USF.
Implementation of Course-Based Learning Communities and Living Learning Communities along with the Development of a Simple Python Program for Measuring Retention

Abstract

Retention of engineering students has been a major concern for universities across the country. Those students who have trouble adjusting and making friends tend to leave at a higher rate than those students who have successfully integrated into college life. In addition, due to the rigor of engineering programs, students have to learn how to study better and more often than they had just a few months prior when they were still in high school.

Learning communities have shown success in helping students to become academically and socially integrated quickly. This has resulted in students being retained higher rates that those students who are not in these types of communities.

This paper chronicles the development and implementation of course-based learning teams and living learning communities at a major metropolitan university, along with the development of a flexible and easy to use python program to measure the retention of students. Many schools do not have robust data systems to provide retention numbers and other data analysis required to measure the success of their programs. By developing these programs in house, in addition to providing basic overall retention numbers, it allows for an expansion of the measurement of retention into several categories and subcategories which helps to provide a more in-depth analysis of student success programs.

Initial findings over the first two years indicate that these learning communities are having a significant impact on the retention of the students not only at the university level but also in the college of engineering. The software program helped to show the basic retention numbers and then allowed for further deeper exploration of student retention by showing the retention broken out by many different subcategories of students.

Introduction

Learning communities have a long history including the Meiklejohn “Experimental College” at the University of Wisconsin in 1920. In the past couple of decades they have emerged as a way to improve the retention for first year students.

During the 1980’s and 1990’s there was a renewed interest in improving undergraduate education in the United States. The Boyer Commission in 1998 released its report, Reinventing Undergraduate Education: A Blueprint for America’s Research Universities, on the state of undergraduate education. It recommended 10 ways to change higher education. Its 10th recommendation states:
Research universities should foster a community of learners. Large universities must find ways to create a sense of place and to help students develop small communities within the larger whole. (p.34)

The Boyer report served as a call to action for colleges to reform their educational practices and restructure classrooms to increase active learning among students.

Scholarly research in the 1980s and 1990s provided the underpinnings of the learning community concept. Vincent Tinto who studied the causes of attrition in college found that students were more likely to stay in college if they connected both academically and socially to the institution. Alexander Astin’s research found that the quality and the quantity of students interactions with peers and faculty were important factors in developing student engagement in the life of the institution.

Lenning and Ebbers (1999) wrote that Alexander Astin and Vincent Tinto models showed the importance of “community” learning and involvement among students and faculty.

The “involvement” model (Astin) and the “student departure” model (Tinto) provide theoretical and conceptual reasons why student learning communities should impact college students positively, and much research supports both models. The models suggest that learning communities should increase students’ development, achievement, and persistence through encouraging the integration of social and academic lives within a college or university and its programs, and through quality interaction with peers, faculty members, and the campus environment. (pp. 49–50)

Making the transition from high school to college can be difficult for many students, resulting in a high rate of attrition in the first year at college. Students who struggle to get integrated into university life are more likely to leave after the first year. The Boyer report states, “There is more of everything…but that complexity can also be baffling and overwhelming to students, making them feel lonely, remote, and too anxious for optimal earning.” (p.34)

Learning communities help students make the sometime difficult transition from high school to college. By the 1990’s, universities across the country were experimenting with various types of learning communities. The learning communities including residential communities can all vary widely in structure and involvement with faculty and staff. Alexander Astin defined learning communities as:

Such communities can be organized along curricular lines, common career interests, avocational interests, residential living areas, and so on. These can be used to build a sense of group identity, cohesiveness, and uniqueness; to encourage continuity and the integration of diverse curricular and co-curricular experiences; and to counteract the isolation that many students feel (Astin, 1985, p. 161).
While learning communities have taken hold across the country at hundreds of universities, there are still places where they have not been introduced beyond the basic residential learning communities.

Course-Based Learning Communities

In 2011, the College started the process to set up course-based learning teams (LTM) linked to first year math courses. For the College, the purpose for creating the LTMs was to improve retention. Sophomore or upper-class engineering students were hired and trained as peer mentors. Mentors were taught how to use cooperative learning techniques to facilitate active learning on the subject matter for an hour. The LTMs ranged in size from 10 students to up to 20 students. The students would meet on campus in a classroom with a peer mentor one hour a week. Additional optional study sessions were also offered throughout the week. In addition, social activities were planned to help the students to get to know each other. To make sure that the one hour a week meeting appeared on the students’ schedules, a zero credit course was created. Students signed up for the LTM session during summer orientation with their advisor’s assistance. Once registered, the LTM course would then reserve the classroom space and show up on the students’ class schedule which reminded students to attend it. The grading for the class was pass/fail.

Engineering Living Learning Communities

The College began a small engineering living learning community (ELLC) of 28 students in the fall of 2007. Two years later, in 2009, the ELLC was then moved to a newer more expensive residence hall in 2009 which had suite style rooms and was located close to the engineering buildings. That year the enrollment more than doubled (see Table 1). Due to limited residential housing there is no more room to expand the ELLC.

Table 1. ELLC Enrollment and Overall Enrollment of First-Year Students

<table>
<thead>
<tr>
<th>Year</th>
<th># Admits</th>
<th>ELLC</th>
<th>ELLC</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007</td>
<td>440</td>
<td>28</td>
<td>6.4%</td>
</tr>
<tr>
<td>2008</td>
<td>479</td>
<td>27</td>
<td>5.6%</td>
</tr>
<tr>
<td>2009</td>
<td>363</td>
<td>71</td>
<td>19.6%</td>
</tr>
<tr>
<td>2010</td>
<td>559</td>
<td>82</td>
<td>14.7%</td>
</tr>
<tr>
<td>2011</td>
<td>499</td>
<td>93</td>
<td>18.6%</td>
</tr>
<tr>
<td>2012</td>
<td>645</td>
<td>100</td>
<td>15.5%</td>
</tr>
<tr>
<td>2013</td>
<td>643</td>
<td>97</td>
<td>15.1%</td>
</tr>
<tr>
<td>2014</td>
<td>694</td>
<td>124</td>
<td>17.9%</td>
</tr>
</tbody>
</table>

Beginning in 2012, live-in peer mentors were hired for each of the residential floors to coordinate social and academic activities for the students. The main goal of the ELLC was to improve the retention rate of engineering students. Social activities were done to promote bonding and group affiliation especially the first couple of weeks since over 75% of students
reported feel lonely in their first two weeks of the semester. Tutoring was also provided in the evening to help students who may be struggling in classes.

Cohort Profiles

Table 2 shows the ethnicity of the ELLC and LTM cohorts compared to the first-year engineering. The 2012 ELLC had slightly fewer Black and Hispanic students, however, the 2013 ELLC, the percentages for Black and Hispanic students was about the same as the overall engineering student population. The LTM had a higher percentage of Hispanic students than the overall population for both 2012 and 2013 at 25% and 23.6% respectively.

Table 2. ELLC and LTM 2012 and 2013 Cohort by Ethnicity

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>2012 ALL</th>
<th>2012 ELLC</th>
<th>2012 LTM</th>
<th>2013 ALL</th>
<th>2013 ELLC</th>
<th>2013 LTM</th>
</tr>
</thead>
<tbody>
<tr>
<td>American Indian or Alaska Native</td>
<td>1.4%</td>
<td>0.0%</td>
<td>1.3%</td>
<td>0.3%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Asian</td>
<td>7.1%</td>
<td>5.2%</td>
<td>3.9%</td>
<td>8.1%</td>
<td>7.3%</td>
<td>9.3%</td>
</tr>
<tr>
<td>Black or African American</td>
<td>7.6%</td>
<td>5.2%</td>
<td>11.8%</td>
<td>9.0%</td>
<td>10.1%</td>
<td>10.7%</td>
</tr>
<tr>
<td>Hispanic or Latino</td>
<td>18.1%</td>
<td>16.5%</td>
<td>25.0%</td>
<td>19.0%</td>
<td>19.3%</td>
<td>23.6%</td>
</tr>
<tr>
<td>Native Hawaiian/Other Pacific Islander</td>
<td>0.3%</td>
<td>0.0%</td>
<td>0.7%</td>
<td>0.5%</td>
<td>0.0%</td>
<td>0.7%</td>
</tr>
<tr>
<td>Non-Resident Alien</td>
<td>5.1%</td>
<td>5.2%</td>
<td>1.3%</td>
<td>6.2%</td>
<td>0.9%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Unknown</td>
<td>1.1%</td>
<td>1.0%</td>
<td>2.0%</td>
<td>0.9%</td>
<td>1.8%</td>
<td>1.4%</td>
</tr>
<tr>
<td>White</td>
<td>59.2%</td>
<td>67.0%</td>
<td>53.9%</td>
<td>56.0%</td>
<td>60.6%</td>
<td>54.3%</td>
</tr>
<tr>
<td>Total Number of Students</td>
<td>645</td>
<td>97</td>
<td>152</td>
<td>643</td>
<td>109</td>
<td>140</td>
</tr>
</tbody>
</table>

Significantly more females enrolled in the ELLC compared to the overall percentage of women in the incoming first year class (Table 3). The LTM for both years mirrored the percentage of first year women overall in the College.

Table 3. ELLC and LTM Cohorts by Gender

<table>
<thead>
<tr>
<th></th>
<th>ELLC</th>
<th>LTM</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>M</td>
<td>N*</td>
</tr>
<tr>
<td>2012</td>
<td>41</td>
<td>56</td>
</tr>
<tr>
<td>2013</td>
<td>36</td>
<td>73</td>
</tr>
<tr>
<td>Overall</td>
<td>291</td>
<td>996</td>
</tr>
</tbody>
</table>

* N – no reported gender.

Retention Data

Since the main purpose of the LTM and the ELLC is to improve the retention rates, the College needed a way to calculate those results. Institutions vary in how their institutional research units are set up to provide this type of data. At this institution, there is not a way to measure retention of learning communities through the existing institutional research unit. Some of the retention reports provided by the university (see Table 4 for the list of reports) yield retention rates at the university level with demographic information and enrollment status (such as ethnicity, gender, and part-time/full-time students) but did not provide retention rates for an individual college.
## Table 4. List of Institutional Retention Reports

<table>
<thead>
<tr>
<th>Title</th>
<th>Institutional Retention Reports</th>
</tr>
</thead>
</table>
| Retention                  | Report displays the enrollment, graduation and retention activity of selected student cohorts. Cohorts vary by selected Report Type (Retention Total University, or IPEDS†, or CSRDE††). At the Campus and College level, students are selected for the cohort by Home Campus and by the college of their first declared major.  
* Limitation: a user may select the cohort’s college declared upon entry to the university, but reports retention rates for the university and not for the college. |
| Term to Term Enrollment    | Report displays the enrollment, graduation and retention activity of selected student cohorts. Cohorts vary by selected Report Type. At the Campus and College level, students are selected for the cohort by Home Campus and by the college of their first declared major, for any benchmark.  
* Limitation: A user may select the beginning term and end term to track a cohort. The report doesn’t indicate percent graduated and percent still enrolled. Also, the report doesn’t specify if attrition is based on lack of enrollment in between the selected beginning and end term, or if it is based on lack of enrollment or graduation for the selected end term. |
| IPEDS Retention Summary    | This report displays a summary of the annual activity of IPEDS cohorts. Data are displayed at the system level with options to split the cohorts by the U Main Campus and the 2<sup>nd</sup> campus (our two campuses with full-time FTIC cohorts).  
* Limitation: the report provides retention for the campus or university only and with predefined cohorts and is not intended for individual colleges. |
| RETTrack                   | This interactive application is to be used to track custom cohorts for retention purposes. It allows users to upload any group of students (identified by UID's) and track them over a 10-year period of time. UID is a student identification number.  
* Limitation: the report does not provide retention or graduation rates for the college, just the university. In addition, it doesn’t provide demographic information. |
| Retention Flow             | This interactive model shows the flow of IPEDS cohort students used for retention to various campuses across a six year period.  
* Limitation: report tracks retention between campuses for predefined cohorts and is not intended for individual colleges. |

†CSRDE (Consortium for Student Retention Data Exchange) – The Consortium for Student Retention Data Exchange at the University of Oklahoma is a consortium of two-year and four-year institutions dedicated to achieving the highest levels of student success through collaboratively sharing data on student’s progress to degree and includes not only First Time in College (FTIC) student tracking but also Transfer student tracking. FTIC cohorts are full-time Fall Term students and transfer cohorts are full-time and part-time Fall Term students.  
[
http://csrde.ou.edu/web/index.html
]

††IPEDS (Integrated Postsecondary Education Data System) - IPEDS is the core postsecondary education data collection program in the U.S. Department of Education's National Center for Education Statistics (NCES). It was designed to help NCES meet its mandate to report full and complete statistics on the condition of postsecondary education in the United States. It is a single, comprehensive data collection system developed to encompass all institutions and organizations whose primary purpose is to provide postsecondary education.
IPEDS is built around a series of interrelated surveys to collect institution-level data in such areas as enrollment, program completions, faculty and staff, and financing.

None of the reports provided the College with retention rates for the learning communities, high school preparation attributes, honors students, first semester performance, or math course, or any number of categories needed to assess retention rates in the College. The only report that gives the retention rate at the college level doesn’t specify percent graduated and percent still enrolled. It also doesn’t specify if attrition is based on lack of enrollment in between the selected beginning and end term, or if it is based on lack of enrollment or graduation for the selected end term and therefore is inaccurate. In addition, the “beginning of term” option is actually then enrollment at the end of the term.

The census date used by the university is further into the semester than is used by AAU institutions. This institution uses a six week benchmark and/or a final benchmark taken after the end of the term. The College needed to use an earlier census date to more accurately reflect enrollment at the beginning of the semester and to be consistent with other institutions. The census date is the time each institution takes a snapshot of their enrollment numbers for reporting each year. All of the benchmark institutions for this university use between the 10th day to the 21st day of the semester.

The College also needed a deeper understanding of retention which required expanding the categories to track students beyond just ethnicity and gender. Due to a lack of institutional reports on retention rates for various engineering student cohorts of interest such as learning teams and living learning communities, conditionally admitted students, honor students, and attributes regarding high school preparation, it became necessary for the College to create a way to produce these reports themselves.

Enrollment data was already being collected by the College using the 10th day for the census date. However, putting this data into the format needed to determine the retention rate was extremely time-consuming as the College wanted to see different breakdowns in the retention rates by categories within categories. But as the categories expanded and additional retention numbers were requested, the time it took to calculate each new cohort was extensive using Excel. If Excel was used, a pivot table would need to be created for each population defined by specific filters and each cohort, since continuation terms and graduation terms vary by cohort. This would result in the need to run multiple pivot tables for a single population to obtain continuation and graduation counts for all cohorts. The number of pivot tables increase exponentially when reviewing different populations for all cohorts. To compare results among all populations and cohorts, the counts from the multiple pivot table runs would also need to be collected in one sheet or table.

The python program offered a much quicker, more efficient, and accurate solution to obtaining the results needed in a view that can be used for analysis. Every time a new category or subcategory wanted to be looked at, it would only involve adding one line to the python code.
Therefore, a python program was created so that additional requests for retention data for new categories or subcategories could be calculated in minutes. An undergraduate computer science major was hired at $10 per hour and spent about 25 hours working on developing the program. The hours include the time the student needed to learn the basics of Python.

Python was chosen since a student group as part of a class project had recently used this language to create what was called a “deficiency” report. This project allowed for reports on students who, for example, failed courses more than the allowed amount, or had a low grade point average for too many semesters in a row, etc. Python is relatively easy to learn, and it is very readable so it is easy to maintain the program. This program used SQLAlchemy which is the Python SQL toolkit to create the databases needed for the reports. (Refer to the Appendix I for details on the program.)

The resulting program allowed the user to know very little about programming but to still be able to go in and add new code to find the retention rate of any new category or subcategory in the database when needed. The program returns the results to an Excel file with each sheet labeled by the main retention category used (see Figure 1).

Figure 1. Output of Retention Program

The initial coding for all the main categories was done by the programmer for the basic retention tables that were needed such as determining the retention rate by gender, ethnicity, honors, ELLC participation, and LTM participation. If the College wanted to know the retention rate of a certain subcategory of students the lines of code that needed to be added were very simple. For example, the following code queries the database and returns the records of students not in the ELLC:

```python
session.query(Student.id).filter(Student.ELLC == 0),
```

If the retention rate of all males in the ELLC was required, a line added to the program would look like the following:

```python
session.query(Student.id).filter(Student.ELLC > 0).filter(Student.gender == 'M'),
```

Or for the retention rate of males who had a high school grade point average of greater than or equal to 3.80, the new inserted code would like the following:

```python
session.query(Student.id).filter(Student.gender == 'M').filter(Student.hsgpa >= 3.8),
```

Refer to Figure 2 for a snapshot of the code that would be entered to find the retention rate of students regarding their participation in the ELLC. This type of capability allowed for an expansive look at retention in the College and an easy way to quickly obtain the retention rates of new subcategories of students without having a programmer available.

```python
#process(student_list, num_tables, num_years_to__run, sheet_num)
#student_list will have the main filter of each table for all students it is session.query(Student.id)
#If you would like to filter by certain attributes that are in the SQL table format is as follows
#session.query(Student.id).filter(Student.ATTRIBUTE_NAME > 0)
#the > 0 is one example of criteria for filtering an attribute that is an integer
#the filter function can be == (equal to), != (not equal to) also
#to filter by more than one just chain .filter(more_things) after

#ELLC
studentList = [session.query(Student.id), session.query(Student.id).filter(Student.ELLC > 0),
               session.query(Student.id).filter(Student.ELLC > 0).filter(Student.hsgpa >= 3.8),
               session.query(Student.id).filter(Student.ELLC > 0).filter(Student.hsgpa < 3.8),
               session.query(Student.id).filter(Student.ELLC == 0),
```

Figure 2. Example of the Programming Code (refer to appendix I for more information)

**Retention Results**

As a result of the Python program, the retention rates could be calculated quickly for different categories of students. The following Figures 1A, 1B, 1C shows the retention rates of students who are in the ELLC, for all students not in the ELLC, and for all students who are not in the Honors learning community, the LTM program, or the ELLC (designated as Non-LLC).

Figure 3A. Persistence to the 2nd year  
Figure 3B. Persistence to the 3rd Year
First year retention rates for students staying in the College who were in the ELLC in year 2013 was 84% compared to 76%. Second year for the 2012 ELLC cohort was 69% to 52% and third retention rate for the 2011 cohort was 60% to 39%. All significant higher for being in the ELLC compared to not being in any learning community. 2010 showed the lowest rates of retention. That year there was a sharp drop in advisors available due to family emergencies. Throughout the 2010-2011 school year the number of advisors available often dropped to as low as two advisors from a high of six full-time advisors.

Table 5. ELLC Students Compared to Non-ELLC Students

<table>
<thead>
<tr>
<th></th>
<th>2012 ELLC students</th>
<th>Non-ELLC students</th>
<th>ELLC Mean</th>
<th>Non-ELLC Mean</th>
<th>F value</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSGPA</td>
<td>97</td>
<td>545</td>
<td>3.88</td>
<td>3.85</td>
<td>0.57</td>
<td>0.4512</td>
</tr>
<tr>
<td>ACTM</td>
<td>69</td>
<td>360</td>
<td>27.38</td>
<td>26.98</td>
<td>0.79</td>
<td>0.3741</td>
</tr>
<tr>
<td>SATM</td>
<td>87</td>
<td>476</td>
<td>631.38</td>
<td>620.67</td>
<td>1.99</td>
<td>0.1591</td>
</tr>
</tbody>
</table>

Table 5 shows the analysis of variance (ANOVA), comparing the 2012 and 2013 ELLC cohorts with those students not in the ELLC. There were no significant difference in high school grade point average (HSGPA), ACT math scores or SAT math scores. HSGPA is the recalculated GPA (see Appendix II). The admissions office recalculates all high school grade point averages to put them on the same scale since the high school GPA varies widely among schools. Fewer schools are reporting high school rank which used to be one of the best ways to evaluate students but now is not reported from about 25% of the students.

Tables 6A and 6B shows the comparison of the ELLC cohorts to all students not in any type of learning community referred to as the non-LLC cohort. 2012 ELLC had statistically higher
means in HSGPA, ACT math and SAT math than the non-LLC students and the 2013 ELLC had a statistically higher ACT math score (ACTM).

Table 6A. Descriptive Statistics Comparing 2012 ELLC Students to Non-LLC Students

<table>
<thead>
<tr>
<th></th>
<th>2012 ELLC Students</th>
<th>Non-LLC† Students</th>
<th>ELLC Mean</th>
<th>Non-LLC† Mean</th>
<th>F Value</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSGPA</td>
<td>97</td>
<td>297</td>
<td>3.88</td>
<td>3.73</td>
<td>10.92</td>
<td>0.0010**</td>
</tr>
<tr>
<td>ACTM</td>
<td>69</td>
<td>193</td>
<td>27.38</td>
<td>25.98</td>
<td>10.49</td>
<td>0.0014**</td>
</tr>
<tr>
<td>SATM</td>
<td>87</td>
<td>261</td>
<td>631.38</td>
<td>605.67</td>
<td>10.79</td>
<td>0.0011**</td>
</tr>
</tbody>
</table>

Table 6B. Descriptive Statistics Comparing 2013 ELLC Students to Non-LLC Students

<table>
<thead>
<tr>
<th></th>
<th>2013 ELLC Students</th>
<th>Non-LLC† Students</th>
<th>ELLC Mean</th>
<th>Non-LLC† Mean</th>
<th>F Value</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSGPA</td>
<td>106</td>
<td>308</td>
<td>3.81</td>
<td>3.74</td>
<td>1.31</td>
<td>0.2534</td>
</tr>
<tr>
<td>ACTM</td>
<td>73</td>
<td>188</td>
<td>27.32</td>
<td>26.20</td>
<td>7.00</td>
<td>0.0086**</td>
</tr>
<tr>
<td>SATM</td>
<td>95</td>
<td>260</td>
<td>611.37</td>
<td>604.69</td>
<td>0.86</td>
<td>0.3550</td>
</tr>
</tbody>
</table>

†Non-LLC: Does Not Include Honors, ELLC, & LTM Students

Course-based Learning Teams

The first year of the learning team program in 2012, participation tilted towards the weaker students. As advisors helped students put together their fall schedule of classes those students who seemed to be borderline proficient in math were encouraged to sign up for the LTMs. Table 7 shows differences in the means of the HSGPA, SATM and ACTM between LTM and non-LTM students. The 2012 LTM cohort had significantly lower scores in all three areas.

Table 7. Comparison of Means between 2012 LTM Cohort versus Non-LTM Students

<table>
<thead>
<tr>
<th></th>
<th>2012 LTM Students</th>
<th>Non-LTM Students</th>
<th>LTM Mean</th>
<th>Non-LTM Mean</th>
<th>F value</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSGPA</td>
<td>152</td>
<td>490</td>
<td>3.75</td>
<td>3.88</td>
<td>11.82</td>
<td>0.0006**</td>
</tr>
<tr>
<td>ACTM</td>
<td>98</td>
<td>331</td>
<td>26.0</td>
<td>27.3</td>
<td>11.45</td>
<td>0.0008**</td>
</tr>
<tr>
<td>SATM</td>
<td>136</td>
<td>427</td>
<td>604.3</td>
<td>628.1</td>
<td>13.97</td>
<td>0.0002**</td>
</tr>
</tbody>
</table>

Note. P**<0.01

From Figure 4 it shows that the 2012 LTM students stayed in engineering at a higher rate than non-LTM students for both their second year (80.9% to 74.4%) and their third year (63.2% to 58.0%) despite having lower test scores and HSGPA.

Figure 4 also shows the retention rate of students who did not participate in any learning community activities such as the honors learning community and the ELLC. These students are referred to as the “Non-LLC” group. An analysis of variance of these students (Table 8) show no significant differences in tests scores or high school grade point average (HSGPA) compared to the LTM students. However, the 1st and 2nd year retention rates were significantly different (80.9% to 69.4% for the 1st year retention and 63.2% to 51.8% for the 2nd year retention.)
Table 8. Comparison of Means between 2012 LTM Cohort versus Non-LLC Students

<table>
<thead>
<tr>
<th></th>
<th>2012 LTM</th>
<th># LTM</th>
<th># Non-LLC</th>
<th>LTM Mean</th>
<th>Non-LLC Mean</th>
<th>F Value</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSGPA</td>
<td>152</td>
<td>297</td>
<td>3.75</td>
<td>3.73</td>
<td>0.24</td>
<td>0.6218</td>
<td></td>
</tr>
<tr>
<td>ACTM</td>
<td>98</td>
<td>193</td>
<td>26.02</td>
<td>25.98</td>
<td>0.01</td>
<td>0.9116</td>
<td></td>
</tr>
<tr>
<td>SATM</td>
<td>136</td>
<td>261</td>
<td>604.34</td>
<td>605.67</td>
<td>0.05</td>
<td>0.8301</td>
<td></td>
</tr>
</tbody>
</table>

Results in Table 9 show that the retention rate of LTM students is significantly higher than that of non-LTM students when controlling for HSGPA.

Table 9. Parameter Estimates of Logistic Regression on Retention

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>SE</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>4.11</td>
<td>2.0008</td>
<td>4.2189</td>
<td>0.0400</td>
</tr>
<tr>
<td>2012 LTM Status</td>
<td>1</td>
<td>-0.39</td>
<td>0.1457</td>
<td>7.1461</td>
<td>0.0075**</td>
</tr>
<tr>
<td>HSGPA†</td>
<td>1</td>
<td>-0.89</td>
<td>0.3475</td>
<td>6.5644</td>
<td>0.0104*</td>
</tr>
<tr>
<td>ACTM</td>
<td>1</td>
<td>-0.01</td>
<td>0.0581</td>
<td>0.0304</td>
<td>0.8615</td>
</tr>
<tr>
<td>SATM</td>
<td>1</td>
<td>-0.002</td>
<td>0.00299</td>
<td>0.2572</td>
<td>0.6120</td>
</tr>
</tbody>
</table>

Note. P**<0.01, p*<0.05.

† HSGPA - The Office of Admissions recalculates students' high school GPA for admission and scholarship purposes. This calculation is based on academic courses and the rigor of the curriculum. This is the HSGPA used here, not the students actual high school GPA. See Appendix II for details.

2013 LTM Results

For the 2013 LTM cohort their ACT and SAT math scores were significantly lower when compared to the Non-LTM group (Table 11A), however, when compared to the Non-LLC group there were no significant differences in HSGPA, ACTM, and SATM (Table 11B).

The retention rate for the 2013 cohort was 76.4%, while for the non-LLC students it was 76.1, which was not significantly different (Table 12).
Table 11A. Comparison of Means between 2013 LTM Cohort versus Non-LTM Students

<table>
<thead>
<tr>
<th></th>
<th>2013 LTM Mean</th>
<th>Non-LTM† Mean</th>
<th>F value</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSGPA</td>
<td>3.83</td>
<td>3.85</td>
<td>0.28</td>
<td>0.5968</td>
</tr>
<tr>
<td>ACTM</td>
<td>26.28</td>
<td>27.16</td>
<td>5.67</td>
<td>0.0177*</td>
</tr>
<tr>
<td>SATM</td>
<td>601.74</td>
<td>621.56</td>
<td>9.59</td>
<td>0.0021*</td>
</tr>
</tbody>
</table>

†Non-LTM group includes ELLC and Honors students

Table 11B. Descriptive Statistics and LTM/Non-LLC Mean Comparison Results

<table>
<thead>
<tr>
<th>Variable</th>
<th># LTM</th>
<th># Non-LLC</th>
<th>LTM Mean</th>
<th>Non-LLC† Mean</th>
<th>F Value</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSGPA</td>
<td>139</td>
<td>308</td>
<td>3.83</td>
<td>3.74</td>
<td>2.69</td>
<td>0.1017</td>
</tr>
<tr>
<td>ACTM</td>
<td>99</td>
<td>188</td>
<td>26.28</td>
<td>26.20</td>
<td>0.05</td>
<td>0.8188</td>
</tr>
<tr>
<td>SATM</td>
<td>121</td>
<td>260</td>
<td>601.74</td>
<td>604.69</td>
<td>0.21</td>
<td>0.6460</td>
</tr>
</tbody>
</table>

†Non-LLC: Does Not Include Honors, ELLC, & LTM Students

Table 12. Chi-square Test on 1st Year Retention for the 2013 Cohort

<table>
<thead>
<tr>
<th></th>
<th># Staying ENGR</th>
<th># Leaving ENGR</th>
<th>% Staying</th>
<th>Chi²</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>LTM</td>
<td>107</td>
<td>33</td>
<td>76.4</td>
<td>0.0075</td>
<td>0.9308</td>
</tr>
<tr>
<td>Non-LLC</td>
<td>235</td>
<td>74</td>
<td>76.1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Because of the Python program, the College could explore the retention rate of students by many different variables. HSGPA became one of the variables that the College was interested in exploring further across all the demographics. Looking at the retention rate for those students in the LTMs and the ELLCs who had HSGPA greater or equal to 3.80 showed (Table 13) significant differences in retention rates to those students whose HSGPA was less than 3.80. While that may have been expected, the retention rate was also higher than that of honors students whose average HSGPA was 4.23.

Table 13. 1st Year Retention Rate of Students by Various Categories

<table>
<thead>
<tr>
<th>All New Students</th>
<th>All Honors</th>
<th>LTM &gt;=3.80</th>
<th>ELLC &gt;=3.80</th>
<th>LTM &lt; 3.80</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year</td>
<td># Admits</td>
<td>% Still Engr</td>
<td># Admits</td>
<td>% Still Engr</td>
</tr>
<tr>
<td>2012</td>
<td>645</td>
<td>76.0</td>
<td>142</td>
<td>83.1</td>
</tr>
<tr>
<td>2013</td>
<td>643</td>
<td>77.6</td>
<td>125</td>
<td>76.0</td>
</tr>
</tbody>
</table>

A HSGPA of 3.80 was chosen for this example although several different HSGPA’s were evaluated by the College since the Python program made it easy to change the GPA. An addition, the HSGPA of 3.80 was the average high school grade point of all graduating seniors over the past six years. LTM and ELLC students who had a HSGPA equal to or greater than 3.80 also showed the highest rates of retention in engineering for the second year of any group including honors students (Table 14).

Figure 4 shows a large difference in retention rates based on HSGPA. For those students in the ELLC who had less than 3.80 HSGPA their retention rate mirrored those of students who were not in the ELLC but had a HSGPA greater or equal to 3.80.
Table 14. 2nd Year Retention Rates By Cohort and High School Grade Point Average

<table>
<thead>
<tr>
<th>Year</th>
<th>ELLC&gt;=3.80</th>
<th>LTM &gt;=3.80</th>
<th>Honors</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>77.4%</td>
<td>75.0%</td>
<td>65.9%</td>
<td>59.2%</td>
</tr>
</tbody>
</table>

Note: Only 10 Honor students had less than a 3.80 GPA or no GPA recorded. Honors students average scores: HSGPA = 4.23, SATM=675.3, ACTM=29.84, 2012=123 honor students, 2013=109 honor students.

Figure 4. 2nd Year Retention Rate, Cohorts Grouped by HSGPA

Hispanic engineering students has the largest population of underrepresented students in the College at 18%. The Tables 15A and 15B show large differences in retention rates among Hispanics engineering students whose HSGPA was greater than or equal to 3.80 compared to those whose HSGPA was less than 3.80.

Table 15A. College of Engineering Hispanic Engineering Retention Rates

<table>
<thead>
<tr>
<th>Year</th>
<th>&gt;=3.8</th>
<th>&lt; 3.8</th>
<th>% Begin 2nd Yr</th>
<th>&gt;=3.8</th>
<th>&lt; 3.8</th>
<th>% Begin 3rd Yr</th>
<th>&gt;=3.8</th>
<th>&lt; 3.8</th>
<th>% Begin 4th Yr</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008</td>
<td>31</td>
<td>46</td>
<td>74.2</td>
<td>69.6</td>
<td></td>
<td>64.5</td>
<td>47.8</td>
<td></td>
<td>54.8</td>
</tr>
<tr>
<td>2009</td>
<td>27</td>
<td>44</td>
<td>81.5</td>
<td>61.4</td>
<td></td>
<td>70.4</td>
<td>40.9</td>
<td></td>
<td>66.7</td>
</tr>
<tr>
<td>2010</td>
<td>41</td>
<td>75</td>
<td>75.6</td>
<td>56.0</td>
<td></td>
<td>56.1</td>
<td>38.7</td>
<td></td>
<td>51.2</td>
</tr>
<tr>
<td>2011</td>
<td>37</td>
<td>58</td>
<td>73.0</td>
<td>51.7</td>
<td></td>
<td>62.2</td>
<td>31.0</td>
<td></td>
<td>62.2</td>
</tr>
<tr>
<td>2012</td>
<td>75</td>
<td>42</td>
<td>88.0</td>
<td>61.9</td>
<td></td>
<td>70.7</td>
<td>35.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td>67</td>
<td>55</td>
<td>80.6</td>
<td>69.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 15B. Retention Rate at the University Level for the Hispanic Engineering Cohort  

<table>
<thead>
<tr>
<th>Hispanic Cohort Size</th>
<th>% Begin 2nd Yr</th>
<th>% Begin 3rd Yr</th>
<th>% Begin 4th Yr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year</td>
<td>&gt;=3.8 &lt; 3.8</td>
<td>&gt;=3.8 &lt; 3.8</td>
<td>&gt;=3.8 &lt; 3.8</td>
</tr>
<tr>
<td>2008</td>
<td>31 46</td>
<td>93.5 82.6</td>
<td>77.4 73.9</td>
</tr>
<tr>
<td>2009</td>
<td>27 44</td>
<td>96.3 77.3</td>
<td>92.6 72.7</td>
</tr>
<tr>
<td>2010</td>
<td>41 75</td>
<td>87.8 78.7</td>
<td>85.4 70.7</td>
</tr>
<tr>
<td>2011</td>
<td>37 58</td>
<td>91.9 67.2</td>
<td>91.9 56.9</td>
</tr>
<tr>
<td>2012</td>
<td>75 42</td>
<td>94.7 85.7</td>
<td>88.0 66.7</td>
</tr>
<tr>
<td>2013</td>
<td>67 55</td>
<td>86.6 80.0</td>
<td>74.2 63.0</td>
</tr>
</tbody>
</table>

Table 16A. Retention Rates in the College of Engineering for White Males  

<table>
<thead>
<tr>
<th>Males, White</th>
<th>% Begin 2nd Yr</th>
<th>% Begin 3rd Yr</th>
<th>% Begin 4th Yr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year</td>
<td>#   #</td>
<td>&gt;=3.8 &lt; 3.8</td>
<td>&gt;=3.8 &lt; 3.8</td>
</tr>
<tr>
<td>2008</td>
<td>152 171</td>
<td>79.6 64.3</td>
<td>57.2 39.2</td>
</tr>
<tr>
<td>2009</td>
<td>119 114</td>
<td>73.9 65.8</td>
<td>65.5 41.2</td>
</tr>
<tr>
<td>2010</td>
<td>136 188</td>
<td>75.0 60.6</td>
<td>60.3 42.6</td>
</tr>
<tr>
<td>2011</td>
<td>158 137</td>
<td>74.7 65.0</td>
<td>66.5 41.6</td>
</tr>
<tr>
<td>2012</td>
<td>221 161</td>
<td>78.7 67.7</td>
<td>63.3 46.6</td>
</tr>
<tr>
<td>2013</td>
<td>204 156</td>
<td>79.9 69.9</td>
<td></td>
</tr>
</tbody>
</table>

While it may seem obvious that those students who come to college with poorer high school grades would be retained at a lower rate, the large difference in the rate is so compelling, that it changes the way one might look at general retention rates. If there is this much difference in retention rates with a cohort based on HSGPA it lends itself to re-questioning the level and type of support services needed to improve retention. In other words, support for students who are only being retained at 36% in engineering might look different if the retention rate reported was 71%; albeit both may be poor but one significantly so. The ability to study retention rates in multiple categories and subcategories helps to clarify the level of support different cohorts of students may need.

Table 16B. White Male Engineering Students Retention Rate at the University  

<table>
<thead>
<tr>
<th>Males, White</th>
<th>% Begin 2nd Yr</th>
<th>% Begin 3rd Yr</th>
<th>% Begin 4th Yr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year</td>
<td>#   #</td>
<td>&gt;=3.8 &lt; 3.8</td>
<td>&gt;=3.8 &lt; 3.8</td>
</tr>
<tr>
<td>2008</td>
<td>152 171</td>
<td>91.4 80.1</td>
<td>82.9 70.2</td>
</tr>
<tr>
<td>2009</td>
<td>119 114</td>
<td>91.6 81.6</td>
<td>89.1 71.9</td>
</tr>
<tr>
<td>2010</td>
<td>136 188</td>
<td>89.0 81.9</td>
<td>83.1 67.6</td>
</tr>
<tr>
<td>2011</td>
<td>158 137</td>
<td>87.3 78.1</td>
<td>82.9 68.6</td>
</tr>
<tr>
<td>2012</td>
<td>221 161</td>
<td>91.0 82.0</td>
<td>87.3 68.9</td>
</tr>
<tr>
<td>2013</td>
<td>204 156</td>
<td>91.2 81.4</td>
<td></td>
</tr>
</tbody>
</table>
Tables 16A and 16B show the large percentage point difference between white males by the HSGPA cutoff of 3.80. (Very few white females in the College have HSGPAs below 3.80.)

Figure 5 shows a graph of the 2nd year retention rates of Hispanics and white male students in the College of Engineering.

![Retention of Students Beginning Their 3rd Year](image)

**Figure 5. 2nd Year Retention Rate, Cohorts Grouped by HSGPA**

**Assessment**

Program assessment is ongoing and will include both qualitative and quantitative methods. Due to self-selection bias with the learning communities it complicates the assessment. In addition, at the time that the learning teams and the enhanced living and learning community began, significant changes were made in the advising of first year students. The additional changes included requiring all new students to attend a semester long orientation course consisting of the following requirements:

- Meeting with an advisor early on (first 3 weeks) to establish a relationship and to handle any current or potential problems
- Learning about campus resources such as tutoring
- Learning about the curriculum requirements and registration for the spring semester
- Provide information regarding stress, time management, study skills for engineering
- Introduced to student engineering organizations clubs
- Introduced to the engineering majors in the college
- Required to attend the career fair to help clarify their goals and interests, to help them learn more about engineering, and to prepare them for internships
- Write a resume for preparation for applying for internships

Another change that began in 2012 which could also have an impact on retention rates required all students not in a department to see an advisor each semester before they could register for
classes. Advising contacts with students increased significantly as a result of this policy change (Table 16). Advising contacts doubled from the year 2010 to the 2013.

Table 16. Advising Contacts with Students

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Sept</td>
<td>307</td>
<td>309</td>
<td>810</td>
<td>948</td>
<td>309%</td>
</tr>
<tr>
<td>Oct</td>
<td>538</td>
<td>712</td>
<td>817</td>
<td>1539</td>
<td>286%</td>
</tr>
<tr>
<td>Nov</td>
<td>1054</td>
<td>1208</td>
<td>1453</td>
<td>1657</td>
<td>157%</td>
</tr>
<tr>
<td>Dec</td>
<td>330</td>
<td>369</td>
<td>534</td>
<td>718</td>
<td>218%</td>
</tr>
<tr>
<td>Total</td>
<td>2229</td>
<td>2598</td>
<td>3614</td>
<td>4862</td>
<td>218%</td>
</tr>
</tbody>
</table>

Written information was collected from the learning teams twice a semester along with a final evaluation. All students were asked to turn in two reflection papers, one at four weeks in and the second one at 11 weeks. The questions asked on the first paper were the following:

- Please write in a few sentences about your biggest challenges to date concerning adjusting to the University.
- Describe the positive experiences you have had and what has helped you to adjust.
- Please describe your comfort level with your course work to date.
- What courses have been the most difficult and what courses do you feel you need further work on to be prepared for the first set of exams? Please explain.
- Have your study strategies been effective? What do you need to be doing differently to get the grade goal you desire? Please explain.
- What study groups, clubs, and social activities are you participating in? How has the learning community helped you get connected with others at the University?

The second set of questions were the following:

- What academic goals did you set for yourself this semester and have you been able to meet them so far?
- What have you found is the key to being successful in problem-solving classes (e.g. math, physics, chemistry)?
- What are a few things that you have learned in the last couple of months about how to study, time management or just school in general that you wished you had known the first week of classes?
- How has being part of a learning community helped you in doing well in classes, meeting other students and getting involved socially, and adjusting to the University and the College of Engineering?
- What strategies/activities have you been able to successfully use to help manage the stress of college life and the multiple demands on your time?
- How can getting to know your peer mentor and instructors be helpful to you in engineering?

One objective of the assignment was to encourage students to reflect on their experience. Research has shown the importance of both reflection and writing. The other reason was to learn about the students’ experiences with their learning team and school so that intervention could begin immediately. The responses also helped in determining the needs of students throughout the semester. A sample of students comments are the following:
“It has help me make new friends who are of my major so we have the same classes and help each other out”
“It has allow me to adjust to college and become more familiar with the college system”
“I have people that I can study with for tests and people I can hangout with”
“…help me realize that I am not the only person lost or confused in some areas of various topics…”
“I really enjoyed the learning community because if one of us had a question on our homework….there was always someone else who had the same questions. If one person knew the answer….they would be able to explain it…”
“I talked to people I never would have in class if it hadn’t been for this learning community”

Conclusions

The data collection is ongoing, however, to date, the impacts of the LTM and ELLCs are positive. Significantly higher retention rates were shown for students in the LTM and ELLC compared to students not in any type of LLC. Even though self-selection bias can cause learning communities to appear as if they are having more of an impact than they really are, the differences in the retention rate are quite compelling. The importance of students quickly finding a peer social group so that they can then focus on academics cannot be overemphasized. By using the Python program, retention rates were calculated by several categories within categories including entry level math courses and grades, learning communities, and other student attributes. Large differences in retention rates within groups became obvious. The additional exploration of retention rates that the Python program allowed, has raised some very interesting questions. This could lead to re-thinking some of the existing support services and perhaps refocusing them to address the higher level of unpreparedness for college among a much larger group of students. Since the ELLC students outperformed all students in retention rates it points to the importance of noncognitive factors such as motivation, commitment to engineering and education, and an interest in social and academic interaction. More work needs to be done to advertise the course-based learning teams to students prior to orientation and in trying to convince university administrators of the importance of providing more space for ELLC students. Currently, honors students receive preference and more space. In addition, improved training of the peer mentors could be done and there could be more faculty and staff involvement with the learning teams.

References


Appendix I

Please contact the author to get the whole program. Below is an outline of the program and snapshots of it.

Retention.py

The first thing that the program does upon executing is to connect to a local database. If that database doesn’t exist it will create a local database in the form of DB or database file. This file is in the form of a SQLite database, which is a form of SQL.

Update Function

The second function that it executes is the update function, which updates the current or newly form DB file. It does this by opening an excel file that using xlrd (xlrd is a library in python that allows python to interact with excel and CSV files) and reading the excel file line by line until it reaches the end. The update function only grabs certain data from each line that is pertinent to the retention such as, when the student was admitted and the college that student was in for each of their terms. Once the function grabs all the needed data from that specific line it is put it in the database, which is organized according to the sql_table.py file. The function will create an entry into the database for every row in the excel file.

Styles Function

The styles function is used to initialize global variables that are used to style the cells in the final excel sheet. This function also is used to define a global excel sheet that is a copy of the template excel workbook. This global copy is what will be written to as the data is processed.

Student List

Student list is initialized with all the queries that will be used to fill each of the tables in the template. Here is a sample of a query:

```
    session.query(Student.id).filter(Student.precalc != 'N').filter(Student.LTM > 0)
```

Session: is a variable that refers to the database connection made in the connect_database() function. This allows the program to access the instance of that connection to the database.

Session. query(Student.id): this is a query of the table that was defined by the Student class in sql_table.py, which is where all the information is stored. This is equivalent to the sql query:

```
    SELECT *
    FROM Student
```

filter(Student.precalc != 'N'):In SQLAlchemy this is equivalent to the WHERE clause in SQL. It just filters the results returned by Session query(Student.id)
SQL equivalent

```python
session.query(Student.id).filter(Student.precalc != 'N').filter(Student.LTM > 0)
SELECT *
FROM Student S
WHERE S.precalc <> 'N' AND S.LTM > 0
```

These queries return tables that correspond to the data that will be used for the tables in each sheet.

**Process**

In each sheet there are multiple tables, each table in segmented into a block and each block is segmented into each year. The process takes information from the block (which year) and the year (each row) combined with the users queries to filters the database to be processed. Now that the relevant information for each row within each block has been determined the processing can begin. Simple calculations to find the number for each cell.

Still in engineering: using the information from the user and the info from that row and block the database is filtered. Once that is done it counts the number of entries that have ‘EN’ for the college for that specific year.

Grad engineering: using the information from the user and the info from that row and block the database is filtered. Once that is done it counts the number of entries that have ‘EN’ under the ‘college code 1’ for that year

% still in engineering: the number of students in engineering for that year the number of grads of that year divided by number of admits for that year.

other coll: using the information from the user and the info from that row and block the database is filtered. Once that is done it counts the number of entries that do not have ‘EN’ for the college for that specific year.

Grad other: using the information from the user and the info from that row and block the database is filtered. Once that is done it counts the number of entries that do not have ‘EN’ under the ‘college code 1’ for that year.

% other: the number of students not in engineering for that year + the number of other grads for that year divided by number of admits for that year

total at USF: still engr + other college + grad_other + grad+_engr

% at USF: total at USF / number of admits for that year

Once it finishes processing that row it moves onto the next row until it finishes the block. It then moves to the next block and so on until it finishes the table and moves onto the next table. Then
the next table until the sheet is done. Once it finishes the sheet it returns to main until the next call.

**sql_table.py**

`sql_table.py` is the file which structures the database. Each class in the python file is equivalent to a SQL table. In the class it defines each column and any keys. In the case of the retention program there is one key and that is the id, which is the students UID. The reason that this was selected for the key is that it is unique to each student and there are no duplicates.

**Screenshots of the program**

```python
from tkinter import askopenfilename, asksaveasfilename, asksaveasfile
from xlrd import open_workbook, XLRDError
from collections import OrderedDict
from xmlutils import copy
from xlwt import *
import tk
import tkMessageBox
import os
import time
import num
import xlrd, xlwt, xmlutils

from sql_table import *
from sqlalchemy import func
from sqlalchemy import distinct
from itertools import imap, chain
from collections import defaultdict
from sqlalchemy.util import KeyedTuple

def update():
    session.query(Student).delete()
    # open excel file and take in data
    try:
        wb = open_workbook("FTIC_Retention_SummerFall_Updated_NoAC.xlsx")
        # Get the needed sheets from xlax.
        ws_data = wb.sheet_by_name('DATA')
        # note data types and if adding filter the 2nd number reference the column
        for curr_row in range(1, ws_data.nrows):
            session.add(Student(id = ws_data.cell_value(curr_row, 0),
                admitted_term = int(ws_data.cell_value(curr_row, 0)[:-4]),
                if int(ws_data.cell_value(curr_row, 0)[5:]) == 8 # to get the catalog year
                else str(int(ws_data.cell_value(curr_row, 0)[:-4]-1)),
                gender = ws_data.cell_value(curr_row, 1),
                race = ws_data.cell_value(curr_row, 13),
                bhsa = 0.0 if not ws_data.cell_value(curr_row, 17) else float(ws_data.cell_value(curr_row, 17)),
                second_term = ws_data.cell_value(curr_row, 31),
                second_colleg = ws_data.cell_value(curr_row, 32),
                third_term = ws_data.cell_value(curr_row, 36),
                third_colleg = ws_data.cell_value(curr_row, 37),
                fourth_term = ws_data.cell_value(curr_row, 41),
                fourth_colleg = ws_data.cell_value(curr_row, 42),
                fith_term = ws_data.cell_value(curr_row, 46),
                fith_colleg = ws_data.cell_value(curr_row, 47),
                sixth_term = ws_data.cell_value(curr_row, 51),
                sixth_colleg = ws_data.cell_value(curr_row, 52),
                term_code_grad = ws_data.cell_value(curr_row, 56),
```
sixth_college = ws_data.cell_value(curr_row,52),
term_code_grad = ws_data.cell_value(curr_row,56),
college_1 = ws_data.cell_value(curr_row,58),
ELLC = 0 if not ws_data.cell_value(curr_row,63) else int(ws_data.cell_value(curr_row,63)),
LTM = 0 if not ws_data.cell_value(curr_row,64) else int(ws_data.cell_value(curr_row,64)),
honors = 'N' if not ws_data.cell_value(curr_row,65) else ws_data.cell_value(curr_row,65),
sss = 'N' if not ws_data.cell_value(curr_row,66) else ws_data.cell_value(curr_row,66),
surrer = 'N' if not ws_data.cell_value(curr_row,67) else ws_data.cell_value(curr_row,67),
fsi = 'N' if not ws_data.cell_value(curr_row,68) else ws_data.cell_value(curr_row,68),
ac = 'N' if not ws_data.cell_value(curr_row,69) else ws_data.cell_value(curr_row,69),
precalc = 'N' if not ws_data.cell_value(curr_row,70) else ws_data.cell_value(curr_row,70),
calci = 'N' if not ws_data.cell_value(curr_row,71) else ws_data.cell_value(curr_row,71))

# Make all changes update to database.
session.commit()
except IOError:
    print("Error!", "File read fail/wrong file type.")
except XLRDError:
    print("Error!", "File read fail/wrong file type.")
def styles():
    this is where all the styles use will be listed

    #examples
    #dataStyle = xlt.easyxf('align: horiz center; border:
    #left thin, top thin, bottom thin, right thin')
    #totalStyle = xlt.easyxf('pattern: pattern solid,
    #fore_colour tan; align: horiz center; border: left thin, top thin, bottom thin, right thin
    #headerStyle = xlt.easyxf('align: horiz center; font: bold True')
    #global top, left, right, none

    top = xlt.easyxf('align: horiz center; border: top thin')
    left = xlt.easyxf('align: horiz center; border: left thin')
    right = xlt.easyxf('align: horiz center; border: right thin')
    none = xlt.easyxf('align: horiz center')

    # copy template book and make global
    # copy template book and make global
    #wb.get_sheet(4).write(18,4,male,left)
    global wb
    global rb
    rb = open_workbook('template.xls',formatting_info=True, on_demand=True)
    wb = copy(rb) # a writable copy

def connect_database():
    global session
    try:
        engine = create_engine('sqlite:///data.db')
    except TypeError:
        messagebox.showinfo("Error", "Unable to create/connect .db file")
    session = create_session(bind=engine, autoflush=False, autocommit=True)
    Base.metadata.create_all(engine)

def process(studentlist, num_tables, num_years_to_run, sheet_num):

    #function list holds filter functions that change from different tables
    #studentList = [session.query(Student.id),session.query(Student.id).filter(Student.ELL > 0)
    #session.query(Student.id).filter(Student.ELL == 0)]
```python
year_data = [2: [Student.second_term, Student.second_college], \
    3: [Student.third_term, Student.third_college], \
    4: [Student.fourth_term, Student.fourth_college], \
    5: [Student.fifth_term, Student.fifth_college], \
    6: [Student.sixth_term, Student.sixth_college]]

# this is the start of the first table to be filled - 5 which is added in the for loop
start_of_table = -1
row = 0

for table in range(0, num_tables):
    student_list = studentList[table]
    start_of_table = 4 + row
    for year in range(2, num_years_to_run + 1):
        # row iterates through each enrollment year
        excel_offset is the offset for the year
        row = start_of_table
        excel_offset = 2 + (year - 2) * 8
        while rb.sheet_by_index(sheet_num).cell_value(row, 0):
            # number admits
            enter = int(rb.sheet_by_index(sheet_num).cell_value(row, 0))
            admits = student_list.filter(Studen.admitted_term == enter).
                count()
            wb.get_sheet(sheet_num).write(row, 1, admits, right)
            # number still engr

            $number still engr$
            still_engr = student_list.filter(Studen.admitted_term == enter).
                filter(year_data[year][0] == ((enter + year) * 100) / 8).
                filter(year_data[year][1] == "EN").
                count()
            wb.get_sheet(sheet_num).write(row, excel_offset, still_engr, left)

            $grad engr$
            grad_engr = student_list.filter(Studen.admitted_term == enter).
                filter(Student.ter..code_grad < ((enter + year) * 100) / 8).
                filter(Student.college == "EN").
                count()
            # subtract all other past grades from this years grades
            add this year grades to the total
            wb.get_sheet(sheet_num).write(row, excel_offset + 1, grad_engr, none)

        $ in engr &
        if admits == 0:
            in_engr_percent = 0
        else:
            in_engr_percent = round((float(still_engr) + float(grad_engr)) * 100 / float(admits), 1)
        in_engr_percent = str(in_engr_percent) + "%"
        wb.get_sheet(sheet_num).write(row, excel_offset + 2, in_engr_percent, none)

        $in other college$
        other_colle = student_list.filter(Studen.admitted_term == enter).
            filter(year_data[year][0] == ((enter + year - 1) * 100) / 8).
            filter(year_data[year][1] == "EN").
            filter(year_data[year][1] == "EN").
            count()
        wb.get_sheet(sheet_num).write(row, excel_offset + 3, other_colle, none)
        # grad other
```
From this point on, researchers inserted lines depending on what they wanted to look for.

Examples:

```python
#grad other
grad_other = student_list.filter(Student.admitted_term == enter).\ 
    filter(Student.tcrn_code_grad < ((enter + year)*100)+8).\ 
    filter(Student.coll_code_1 != "EN"),\ 
    filter(Student.coll_code_1 != ""),\ 
    count()

#subtract all other past grads from this years grades
wb.get_sheet(sheet_num).write(row,excel_offset+4,grad_other,None)

# in other %
if admits == 0:
    in_other_percent = 0
else:
    in_other_percent = (round((float(other_coll)+float(grad_other))*100/float(admits)),1)

wb.get_sheet(sheet_num).write(row,excel_offset+5,in_other_percent,none)

#st USF
totalAtUSF = still_engr + other_coll + grad_other + grad_engr
wb.get_sheet(sheet_num).write(row,excel_offset+6,totalAtUSF,none)

#percent at USF
if admits == 0:
    atUSF_percent = 0
else:
    atUSF_percent = (round((float(totalAtUSF)*100/float(admits)),1))
atUSF_percent = scr(atUSF_percent) + "%"
wb.get_sheet(sheet_num).write(row,excel_offset+7,atUSF_percent,right)

row += 1

print " sheet ", sheet_num, " is finished"

if __name__ == "__main__":
    connect_database()
    update()
    styles()
    #process(student list, num_tables, num_years_to _run, sheet_num)
```

From this point on, researchers inserted lines depending on what they wanted to look for.

Examples:
```python
# ELLC
studentList = [session.query(Student.id).filter(Student.ELLC > 0).order_by(Student.id.asc())]

# LTH
studentList = [session.query(Student.id).filter(Student.LTH > 0).order_by(Student.id.asc())]

# Gender
studentList = [session.query(Student.id).filter(Student.gender == 'Y').order_by(Student.id.asc())]

# Race
studentList = [session.query(Student.id).filter(Student.race == 'Y').order_by(Student.id.asc())]

# Math
passingGrade = ['A', 'A-', 'A+', 'B-', 'B+', 'C-', 'C']

# not used
studentList = [session.query(Student.id).filter(Student.math < 3.8).order_by(Student.id.asc())]

# LTH and not LTH for precalc and calc
session.query(Student.id).filter(Student.math >= 3.8).order_by(Student.id.asc())

# ELLC and not ELLC for precalc and calc
session.query(Student.id).filter(Student.math < 3.8).order_by(Student.id.asc())

# LTH and ELLC and honors or not LTH and not ELLC and honors for precalc and calc
session.query(Student.id).filter(Student.math >= 3.8).order_by(Student.id.asc())

# LTH and not LTH for precalc and calc
session.query(Student.id).filter(Student.math < 3.8).order_by(Student.id.asc())

# save the workbook
wb.save('Counts Mary.xls')
```
Python code for the SQL table.

```python
from sqlalchemy import Column, Integer, String, ForeignKey, create_engine, and_, or_
from sqlalchemy.orm import relationship, create_session
from sqlalchemy.ext.declarative import declarative_base

Base = declarative_base()

class Student(Base):
    ""
    holds all the data
    ""
    __tablename__ = 'student'
    id = Column(String, primary_key=True)  # the key must be unique
    race = Column(String)
    gender = Column(String)
    admitted_term = Column(Integer)
    admitted_college = Column(String)
    second_term = Column(Integer)
    second_college = Column(String)
    third_term = Column(Integer)
    third_college = Column(String)
    fourth_term = Column(Integer)
    fourth_college = Column(String)
    fifth_term = Column(Integer)
    fifth_college = Column(String)
    sixth_term = Column(Integer)
    sixth_college = Column(String)
    term_code_grad = Column(Integer)
    coll_code_1 = Column(String)
    RILC = Column(Integer)
    LIM = Column(Integer)
    honors = Column(String)
    ass = Column(String)
    summer = Column(String)
    hasgpa = Column(Integer)
    rst = Column(String)
    ec = Column(String)
    precalc = Column(String)
    calc = Column(String)
```
High School GPA Recalculation

High school GPA is recalculated based on grades earned in high school only in core academic subject areas, as well as specified AP and IB fine and performing arts courses. The below table shows quality points added for approved AP, IB, AICE, Honors and Dual Enrollment courses provided a "C" or better is earned.

Table 1. Recalculated High School Grade Point Average*

<table>
<thead>
<tr>
<th>Course Type</th>
<th>Quality Point</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advanced Placement</td>
<td>1.0</td>
</tr>
<tr>
<td>International Baccalaureate</td>
<td>1.0</td>
</tr>
<tr>
<td>Dual Enrollment</td>
<td>1.0</td>
</tr>
<tr>
<td>AICE</td>
<td>1.0</td>
</tr>
<tr>
<td>Honors</td>
<td>0.5</td>
</tr>
</tbody>
</table>

*Source - [http://www.usf.edu/admissions/freshman/app-requirements/gpa-test-requirements.aspx](http://www.usf.edu/admissions/freshman/app-requirements/gpa-test-requirements.aspx)

The purpose is to try to create the same scale for all students as high school GPA varies widely among schools. Fewer schools are reporting high school rank which used to be used as one of the main way to evaluate students but now is now missing from 25% of the students from Florida schools.