Graduation and Retention Study
Phase 1
Basic Transition Analysis

Graduation rates have become an important measure of success for colleges and universities, especially those with selective admissions. When admissions standards are high and many of the applicants are not admitted, most people conclude that every freshman has the ability to graduate. Of course, changes that occur in students' lives can significantly alter the course of their education over four years. However, the general public expects that a highly ranked university will have a high graduation rate, and many of the best universities do have excellent graduation rates.

Consequently, universities regularly examine their own graduation rates and explore routes toward improving them. In this arena, casting blame in many directions is easy, but proving a particular cause and effect relationship is hard. We intend to take a road less traveled by analyzing in more detail than usual the student enrollment behavior over six years. That analysis will identify a set of governing metrics, which the campus can use to collectively work on improving the graduation rates, with predictable results.

Historically the freshman retention rate (percent of freshman class that returns in the fall of their second year) has been considered the key driver for graduation rates. Although Lehigh has been working to improve freshman retention for this reason, the record actually shows stability. As part of this project we have built very accurate files of the freshman cohorts from 1994 to 2000. For the 1994 to 1997 cohorts, the data used for the initial phase of this project, the combined freshman retention rate was 93.1% and there were no statistically significant changes during those years.

However, among Lehigh's peers 93% freshman retention rates predicts a much higher six-year graduation rate than Lehigh has had. For example, a simple regression on last year's peer data predicted an 87% graduation rate for a freshman retention rate of 93%. Lehigh's graduation rate was down to 80% for the 1993 cohort and for the last two years has been only averaging between 83 and 84%. Why? Finding the answer is the prime motivation behind this study. It will not refute the importance of freshman retention but it will challenge its premier status at least at Lehigh.

The model used is a probabilistic one based on recent data and can rigorously predict graduation rates from retention goals. It is constructed by capturing and tracking entire freshman cohorts using the new Banner system, which was not feasible until this spring. In Phase 1 of this study we have examined how students move through three states - active, inactive and graduated. The

students were tracked from fall semester to fall semester, but the same tracking could be done for fall and spring semesters.

For each transition, that is, first year to second year, second year to third year, etc., the frequency at which the students moved from one state to another was determined from the data. For example, in the 1994 cohort 91% of this cohort that earned grades for courses in the fall of their second (sophomore year) year also earned grades for courses in the fall of their third year (junior year). Since 0% of them graduated before the their junior year, 9% did not successfully complete any courses in the fall of their junior year. Thus from the second to third year there was a 91% transition frequency from active to active and a 9% transition frequency from active to inactive for the 1994 cohort.

Altogether the model contains 56 such transition frequencies, and almost half of them are known to be 0 or 1. For example, the transition frequency from graduated to active or inactive must logically be 0 to make the model work. However, there are other less obvious ones that play a role. The transition frequency from inactive to active was 14% between the sophomore to junior year for the 94-cohort and the transition frequency from active to graduated was 1% between the junior and senior year. For the later years there is also a positive inactive-to-graduated frequency. Although some of these frequencies are small, keeping track of all of them is essential to make the model work.

In the analysis we treat these frequencies as transition probabilities and examine several basic questions. First, are there statistically significant differences in these probabilities from one year to the next? Can we detect any trends or are we just seeing fluctuations like those in the number of heads generated from tossing a coin 1100 times? Second, what effect will changes in a specific transition probability have on the graduation rate? For example, assuming no secondary effects, what impact does a one percent increase in the sophomore to junior retention (active to active transition frequency from second to third year) have on the six-year graduation rate? Third, what is the optimal path toward higher graduation rates? In particular, what retention goals would produce an 88% graduation rate?

This model has five important variables, three major and two minor ones. The major ones are the active-to-active transitions between the freshman and sophomore year (the usual freshman retention rate), the sophomore to junior year, and the junior to senior year. For convenience we will refer to them as freshman, sophomore, and junior retention. The minor ones are the inactive-to-active between both the sophomore and junior years and the junior and senior years.

We have the data for these variables from the 1994, 1995, 1996, and 1997 cohorts. Using this data we tested the hypothesis that the frequencies for different years represented unequal underlying probabilities in each of the five

variables. No statistically significant changes appeared in the major variables over these 4 cohorts. In fact, the changes did not even come close to being significant. The inactive-to-active frequencies are less stable and show significant changes in the sophomore to junior transition. However, the numbers are small (2 to 12 out of 61 to 74) and erratic and a pooled sample is probably more informative. From this analysis we concluded that we should combine the data from the four cohorts into a single model with data for 4300 students in the first three transitions.

The nature of the model is such that when a single variable changes the graduation rate changes linearly, and we can calculate the rate or slope of this change in the graduation rate. Calculating these slopes produce the first surprises. The slopes of the graduation rate with respect to freshman, sophomore, and junior retention are 0.41, 0.46 and 0.54 respectively. Thus for each one percent increase in freshman retention we gain 0.41% increase in the six-year graduation rate assuming all the other transition frequencies remain unchanged. Consequently, Lehigh is in a position where improvements in junior and sophomore retention will have a bigger impact on the graduation rate than will increased efforts on the freshman retention rate. In fact, Lehigh cannot hope to have an 88% six-year graduation rate by increasing only the freshman retention rate. The other retention rates must increase as well.

Finally, the model can help set goals for all five of the important variables. The optimal route for increasing the value of a multi-variable function is to follow the gradient, which is easy to do for the six-year graduation rate. Of course, one can not hope to control retention percentages to several decimals, but the numbers do provide an order of magnitude that can be used to set measurable goals with predictable results. The following table constructed from the four cohorts using the model shows the rates that are the results of optimal changes in each of the five variables to produce an 88% graduation rate and sample goals based on them that would also produce an 88% graduation rate:

Optimal Route Rate	**Model Value**	**Sample Goals**
Freshman Retention	95.5%	95.0%
Sophomore Retention	94.2%	95.0%
Junior Retention	98.4%	97.5%
Inactive Sophomore Return	9.5%	10.0%
Inactive Junior Return	18.3%	20.0%
Graduation Rate	88.2%	88.0%

Phase I Findings

1. If no inactive students return for their junior or senior years, the six-year graduation rate would fall by 2%. Consequently, staying in touch for a semester or two with any student who becomes inactive would help maintain and modestly improve Lehigh's current 84% six-year graduation rate.

2. Lehigh cannot achieve an 88% graduation rate just by improving freshman retention. An impossible 99% freshman retention would yield at most an 86.5% graduation rate without improvements in sophomore and junior retention rates. Consequently, we must address retention more broadly in order to increase graduation rates.

3. The retention rate (fall to fall) for active sophomores has actually been lower than the freshman retention rate for 4 straight years. Moreover, increasing freshman and junior retention to 95% and 97% respectively and not increasing sophomore retention yields less than an 86% six-year graduation rate. Consequently, in any efforts to improve graduation rates we must give high priority to understanding why sophomores do not return for their junior year.

4. Even though the junior retention rate is the best of the three, increasing it has the biggest payoff. A 1% increase in junior retention produces a 0.54% increase in the six-year graduation rate in the model while similar increases in freshman and sophomore retention produce only 0.41% and 0.46% increases respectively. Consequently, improving junior to senior retention would contribute noticeably to increasing graduation rates.

5. Given proportional improvements in retention, the return of inactive students, and the resulting increases in the six-year graduation rate, the model consistently predicts even greater increases in Lehigh's four-year graduation rate. Consequently, there are secondary benefits to broadly improving retention with the goal of improving our six-year graduation rate.

Next Steps

- Develop a systematic and informative telephone interview process for all non-returning students this fall and use it to gain a behavioral understanding of the retention rates.

- Understand the role the college enrollment plays in retention by refining the model to track college transitions along with the three enrollment states.

- Use data mining techniques to look for key indicators for failure to graduate.

- Understand the role that the grade point average plays in retention by refining the model to track changes in grade point average along with the three enrollment states.

- Link the results back to admissions data to help identify the characteristics of applicants most likely to earn their Lehigh degree as students.