I give permission for public access to my Honors paper and for any copying or digitization to be done at the discretion of the College Archivist and/or the College Librarian.

Signed_____

Caleb Boyd Standafer III

Date _____

Pre-enrollment Factors Affecting Retention at Rhodes College

Caleb Boyd Standafer III

Department of Economics and Business Administration
Rhodes College
Memphis, Tennessee

2008

Submitted in partial fulfillment of the requirements for the
Bachelor of Arts degree with Honors in Economics

This Honors paper by Caleb Boyd Standafer III has been read and approved for Honors in the Department of Economics and Business Administration.

Dr. Marshall K. Gramm
Project Sponsor

_____

Dr. C. Nicholas McKinney
Second Reader

_____

Mr. James Eckles
Extra-Departmental Reader

_____

Dr. Deborah N. Pittman
Department Chair

_____

CONTENTS

LIST OF FIGURES

ABSTRACT

Pre-enrollment Factors Affecting Retention at Rhodes College

by

Caleb Boyd Standafer III

This paper seeks to use both the theoretical and empirical techniques of economics to build upon the literature of the factors responsible for retention at colleges and universities, and apply it specifically to an area that has not been examined in the literature so far, a small, highly-selective liberal arts college. Rhodes College is the target institution, with the data coming from its extensive databases on student characteristics and biographical and demographic information. Using both an institutional research theoretical model specific to college retention that is seen in much of the literature on the subject, a signaling model from economic literature is added to provide a framework under which incentives can be structured so that certain pre-enrollment characteristics show an increased propensity to stay at Rhodes College beyond a student's freshman year. A probit model is then used to estimate the exact characteristics that identify the increased propensity for a retained outcome, including both academic and social factors. In light of the results, specific policy implications are discussed.

Introduction

This paper looks at the issue of freshman retention, also known by its opposite attrition, at Rhodes College. Freshman retention is a key metric in higher education because while graduation is the ultimate goal, the majority of students who leave a school do so between their freshman and sophomore years. In my work as an Institutional Research Analyst in the Information Services department through the Rhodes Student Associate Program, I have had access to the extensive databases the college has. The data contained in them is ideal for investigating many research questions related to the economics of higher education. I hope to use the game theoretic concept of signaling to provide the theoretical framework to search out with econometric methods signals that students send that separate them into groups that are highly likely to persist at Rhodes from their freshman to their sophomore year, or to attrit.

Rhodes has extensive datasets in digital database form going all the way back to 1987. These databases consist of just about all of the information the school knows about a student, including academic, financial, demographic, and extracurricular information. The school also has data from the administration for the past 25 years of the Cooperative Institutional Research Program (CIRP) Freshman Survey from the Higher Education Research Institute (HERI) at the University of California, Los Angeles (UCLA). This data provides a rich source of information on individual characteristics that is difficult to find elsewhere and is relatively homogenous in terms of its completeness and depth across time. The CIRP Freshman Survey, in particular, provides a lot of detailed demographic information.

The type of information that Rhodes has available to me, with its breadth and depth, does not appear in the literature to have been used to investigate the question of retention before. The type of detailed information in the CIRP Freshman Survey especially stands out as not having been used before. The game theoretic approach is also missing. Additionally, the specific case of retention at a small, selective liberal arts college is absent from the literature.

Marcus (1989) did a study using data from published college guidebooks to look at the factors affecting retention rates at a sample of private four-year colleges and universities in the United States. He found that "the acceptance rate, a required interview, and the average collegiate SAT score" are important factors in determining the retention rate of private colleges in the U.S. These findings suggest that looking at the issue of retention of the institutional level at Rhodes would be beneficial, as all three factors that Dr. Marcus identifies suggest that being more discriminating in admitting students based on a number of factors is a way to increase retention rates. This makes the results of the proposed research important not just for academic reasons, but for policy-making decisions as well. The categories of SAT, interview, and acceptance rate thus provide the base for a theoretical model to build the research off.

Another study that looked at retention was done by Wetzel, O'Toole, and Peterson (1999). They used as their theoretical framework a model developed by Tinto (1975) that characterizes a student's probability of persisting as a function of goal commitment and institutional commitment, these being the dedication a student has to achieving a certain result and their loyalty to a certain academic institution, respectively. They found academic and social integration factors to be the most important factors

affecting retention. Financial considerations, however, were not as important. This could change in research done at Rhodes, however, as Wetzel et al. used "an urban public university enrolling large numbers of non-traditional students" as their population, a different population category to be sure than Rhodes. The framework, however, of looking at academic and social factors, respectively, and determining the role financial considerations play in the commitment a student has to the institution seem to be valid and promising ones.

In looking at retention, I hope to use a signaling model to show that entering freshman have certain characteristics that signal their intention to transfer or to persist on to their sophomore year at Rhodes. This idea came from the concept of Spence (1973) Signalling, which won Spence the Noble Prize in Economics in 2001. The model shows that theoretically, education can act as a signal in the job market regardless of any actual gains in productivity by the worker, simply because it shows that the worker has high aptitude.[1]

Obviously the signaling literature needs to be looked at further, as Rhodes is not offering an incentive wage to every student, though it does to some, such as the Bellingrath Fellows. What I hope to find, however, is that there is a real, experimentally confirmable separating equilibrium for incoming students based on certain personal characteristics that can be described as signals. These signals can be used then to identify students who will persist at Rhodes from their freshman to their sophomore year. These findings can then hopefully be used by other researchers to replicate the study at other schools to find the signals that are relevant for their institutions, and also to do a study of a number of schools together to find signals that are relevant across institutions.

[1] Thank you to Prof. Daniel Arce for introducing me to this concept.

Literature Review

The major theoretical model for undergraduate higher education retention was developed by Tinto (1975) [see Fig. 1]. Tinto believed that retention in post-secondary education was a function of three factors: goal commitment, institutional commitment, and individual characteristics. Goal commitment is a student's determination to accomplish a certain purpose, namely, graduation. Institutional commitment is a student's attachment to a particular college or university as the place to accomplish their goal. Individual characteristics are the things that make every student unique.

Tinto's Model

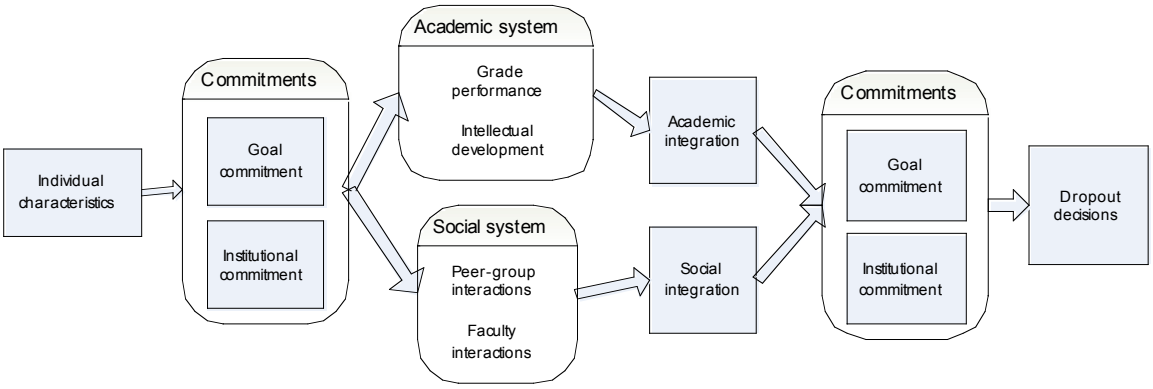

Fig. 1[2]

The basic premise of Tinto's model is that a student must be sufficiently integrated into either the academic or social life, or some combination of both, of the institution for them to decide to stay at the institution. Goal commitment, institutional commitment, and individual characteristics contribute to the initial academic and social integration. This integration in turn then reinforces the goal and institutional commitments. A committed student is then more likely to be retained. This model is useful in that it helps identify certain characteristics, such as family background, pre-

[2] I am grateful to Jay Eckles of Rhodes College for this diagram.

college education, etc., that can be targeted in identifying students who will have a strong commitment to staying at an institution through graduation.

This has been investigated by Braxton, Sullivan, and Johnson (1997), who found that the theory is flawed in that a minority of its propositions are actually testable hypotheses. Despite these shortcomings, there has been no major alternative offered in the literature. Sadler, Cohen, and Kockesen (1997) did develop a model according to Tinto's theory. They emphasize, however, that the independent variables that affect retention are very specific to each institution. Montmarquette, Mahseredjian, and Houle (2001) also developed a model that was successful and consistent with Tinto's theory.

In a study of institutions of higher education in the United Kingdom, Longden (2006) takes on the issue of the decrease in student retention rates that came along with the increase in access to higher education in the UK. He is most interested in the issue of whether students with a high probability of being retained should be recruited or the institution should mold itself to be a place students want to stay at. He cites York and Longden (2004) on four proposed reasons why students leave an institution of higher education:

> "1. flawed decision-making about entering the programme;
> 2. students' experience of the programme and the institution generally;
> 3. failure to cope with the demand of the programme; and
> 4. events that impact on students' lives outside the institution."

The decision to leave is most often based on some combination of the four reasons. These reasons are then used as a framework to look cursorily at trends in the data. Longden's main conclusion is that students must feel integrated into the institution in

order to make a decision in favor of retention. He echoes Tinto's (2002) conclusion that using an "add a course" strategy of remedial aid for struggling students in the form of first-year seminars or mentoring programs, may actually do more harm than good. There is no evidence provided for this belief, however.

In economics, issues of retention have been investigated in higher education in a study by Wetzel, O'Toole, and Peterson (1999). They used as their theoretical framework the model developed by Tinto that characterizes a student's probability of persisting as a function of goal commitment and institutional commitment, these being the dedication a student has to achieving a certain result and their loyalty to a certain academic institution, respectively. They found academic and social integration factors to be the most important factors affecting retention. Financial considerations, however, were not as important. This could change in research done at Rhodes, however, as Wetzel et al. used "an urban public university enrolling large numbers of non-traditional students" as their population, a different population category to be sure than Rhodes. The framework, however, of looking at academic and social factors, respectively, and determining the role financial considerations play in the commitment a student has to the institution seem to be valid and promising ones.

Marcus (1989) did a study using data from published college guidebooks to look at the factors affecting retention rates at a sample of private four-year colleges and universities in the United States. He found that "the acceptance rate, a required interview, and the average collegiate SAT score" are important factors in determining the retention rate of private colleges in the U.S. These findings suggest that looking at the issue of retention of the institutional level at Rhodes would be beneficial, as all three

factors that Dr. Marcus identifies suggest that being more discriminating in admitting students based on a number of factors is a way to increase retention rates. This makes the results of the proposed research important not just for academic reasons, but for policy-making decisions as well. The categories of SAT, interview, and acceptance rate thus provide the base for a theoretical model to build the research off.

Hotchkiss, Moore, and Pitts (2006) evaluate the success of freshman learning communities in increasing retention at an urban, largely commuter campus. The communities are organized around a central theme, and the students in them take the same classes and participate in extracurricular activities together. The authors seek to correct for self-selection into the learning communities. The learning communities were found to have a fairly dramatic effect on the GPA of its members, while also have a moderate positive effect on retention.

Stock, Finegan, and Siegfried (2006) investigate the retention of PhD students in economics graduate programs. Factors that were found to increase retention were being a top 15 program, having students with both higher verbal and quantitative GRE scores, and students having research assistantships. Poor academic performance was the strongest motivator of attrition, and transfer students accounted for a small number of those who left their programs.

Heath (1993) holds that students choose a college or university based at least in part on the local and global status that a degree from that institution will provide. This affects later retention decisions. Stratton, Wetzel, and O'Toole (2005) use a national sample show that the timeframe under which retention is defined significantly affects the results of the analysis of the factors affecting the retention rate. Singell (2004) finds that

need- and merit-based financial aid at a large state university had positive effects on retention, though the precise effects were clouded by bias due to selection effects and the variation of need and ability of the recipients. Desjardins, Ahlburg, and McCall (2002) showed that the structure of financial aid packages has a significant effect on retention beyond simply the absolute dollar amount.

Dearden, Emmerson, Frayne, and Meghir (2005) show that need-based grants given to students undergoing post-compulsory education in England has a positive effect on retention. Angrist, Lang, and Oreopoulos (2006) experimented with methods for reducing high attrition, delayed completion, and poor achievement in first-year undergraduates at a large Canadian university. They found that using a method that combined "peer advising and organized study group services" with "substantial merit-scholarships for solid, but not necessarily top, first year grades" increased retention more than either of the methods alone.

Sacerdote (2000) found in a study of Dartmouth College students that roommates have a significant effect on the academic effort a student puts forth and therefore their subsequent GPA, as well as the decision to become socially integrated in fraternity, where the effects include not just roommates but also the whole dorm. These effects do not seem to affect other decisions made by freshman. Interestingly, social integration due to roommate effects increased the likelihood of roommates staying together the next year, while academic integration due to roommate effects decreased the likelihood of the roommates staying together for sophomore year.

Johnes and McNabb (2004) examine retention in the United Kingdom, focusing on the reasons for both voluntary and non-voluntary dropout. The match between a

student and an institution of higher education is found to be important, with a student who is the right 'fit' for an institution and vice versa has a much stronger chance of being retained. Of importance as well is the effect of peer groups on retention. An especially interesting and unique finding is that "universities that have high standards of quality in learning and teaching have lower dropout rates than others that do not achieve the same standards." Their research confirms Sacerdote's findings, demonstrating that matching and peer group effects are very important to the question of whether a student is retained.

Mixon and Trevino (2005) demonstrated that a university's football success is positively correlated with both freshman retention and graduation rates. Langbein and Snider (1999) demonstrate that enrollment in courses consistently rated as either poor or top increases the probability of attrition, providing a look at the effect of teacher quality on retention. Robst, Keil, and Russo (1998) found that the retention of female students increases as the number of their mathematics and science courses taught by female faculty members increases. The effect is more pronounced when the proportion of female students in an individual class is small. Levy and Murray (2005) found that "an appropriately supportive transitional program and environment" increases the retention rate of students who normally would not qualify for a college education.

Huang, Lin, and Chuang (2006) used a mixture of human capital theory, signaling models, and regression analyses to evaluate worker retention at an individual firm. They found that many factors affected the retention levels, including firm-specific human capital, wages, and signaling effects. The affects of firm-based factors on retention decisions were much greater than individual factors.

There is also significant literature on the issue of military retention. Fullerton (2003) found that the primary reason for pilot attrition in the Air Force was the pay gap between private airlines and the United States Air Force. This was in contrast to the stated claim that he was investigating of whether the high operations tempo was the primary reason for pilots leaving.

Daula and Moffitt (1995) investigated the incentives for military reenlistment. They found that the military-civilian pay differential is a major factor in military reenlistment decisions, just as Fullerton found, and specifically that differential at the time the reenlistment decisions is made.

Theory

Spence (1973) provides a signaling model for employers seeking to hire high-ability employees. The model holds that an employer can gauge the ability level of a worker by the amount of education they receive. This stems from the fact that education is more costly to lower-ability workers than to higher-ability ones, due to time, effort, etc. Thus, it is only worthwhile for high-ability workers to become highly educated. Therefore, education acts a signal, separating high-ability workers from low-ability ones.

The following inequalities apply to the game [see Fig. 2[3]]:

$a_H > a_L$; $w_C > w_o$; $s_H > s_L$

The game begins with a move by nature that chooses an individual's ability level, or to phrase it differently their ability to commitment to being retained, as either high (a_H) with probability *q* or low (a_L) with probability *1-q*. From there, each player, knowing her own type, makes a decision about whether to receive post-secondary education (s_H) or forego

[3] I am grateful to Prof. Dan Arce for this representation of Spence's model.

it (s_L). The next move is made by employers, who have to make a decision about what wage to offer a prospective employee whose aptitude is unknown to them. They can

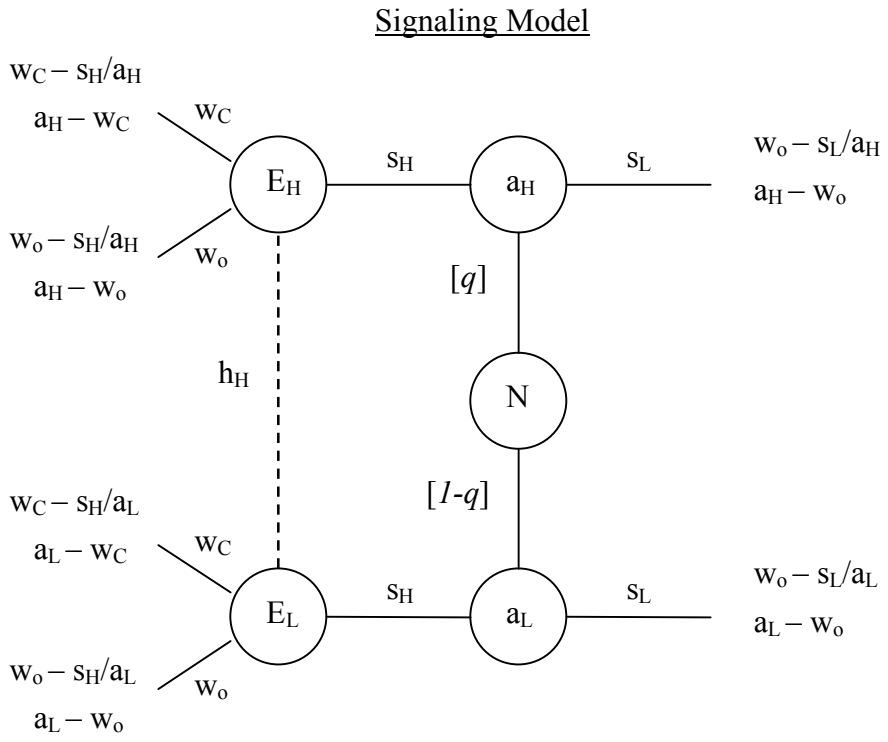

Fig. 2

either offer a competitive wage (w_C) or an ordinary wage (w_o). The payoff to them is either the high or low aptitude output of the employee minus the wage they pay. The employee's payoff is the wage they are paid minus the cost of their educational decision, defined as their educational decision divided by their aptitude level. Based on the inequalities defined above, this means that it is less costly for high aptitude types to obtain education than it is for low aptitude types. The game can be truncated as shown on the s_H side because if the low education signal is being sent, it only makes sense for the employer to offer w_o, which is born out by the mathematics using the same process used below.

In investigating the equilibria, it is instructive to look at two separate cases: pooling and separating equilibria. The relevant pooling equilibrium is one in which both high and low aptitude types receive a college education. This equilibrium is important because it shows whether an employer ever has to worry that it will hire a low aptitude employee who sends the education signal. The relevant separating equilibrium is one in which the high aptitude types receive a college education and the low aptitude types do not. This demonstrates whether the employer can reliably consider a college education a signal of high aptitude.

To investigate the separating equilibrium where a_H sends s_H and a_L sends s_L, we first must now the employer's belief that it is at E_H. This is given by Bayesian updating through

$$\mu_E(E_H|s_H) = [\text{prob}(a_H) \text{ x prob}(a_H \text{ sends } s_H)] / [\text{prob}(a_H)\text{xprob}(a_H \text{ sends } s_H) + \text{prob}(a_L) \text{ x prob}(a_L \text{ sends } s_H)]$$

This equals

$$\mu_E(E_H|s_H) = q(1)/[q(1) + (1-q)(0)] = 1$$

which means that the employer can believe that it is certain that if the worker is sending s_H, she is a high-aptitude type. It is a best-reply for the employer to offer w_C, then, if the worker is sending s_H.

To see whether (s_H, w_C) to be an equilibrium, s_H must then also be the high aptitude worker's best reply to w_C. In order for a_H to send s_H,

$$w_C - s_H/a_H \geq w_o - s_H/a_H$$

must be satisfied. This simplifies to

$$a_H(w_C - w_o) \geq (s_H - s_L)$$

which means that the bonus (w_C-w_o) for having high aptitude must be greater than the cost of high education.

Likewise, in order for a_L to send s_L,

w_o-$s_L/a_L \geq w_o$-s_H/a_L

must be satisfied. This simplifies to

$(s_H$-$s_L) \geq (w_C$-$w_o)a_L$

which means that the bonus (w_C-w_o) for having high aptitude must be greater than the cost of high education. It has already been established intuitively that w_o is the employer's best reply to s_L. Combining the results for a_H and a_L results in the following inequality

$a_H(w_C$-$w_o) \geq (s_H$-$s_L) \geq a_L(w_C$-$w_o)$

Thus, the conditions of the separating equilibrium are satisfied as (s_H, w_C) and (s_L, w_o) are mutual best replies, but only if the bonus being offered to high aptitude workers is at least as great as the cost to them of their education, but still not enough to overcome the cost of education to the low aptitude types.

To test the relevant pooling equilibrium where both aptitude types send s_H, we must again start with Bayesian updating to determine the employer's belief that they are at E_H if the worker is sending s_H. This results in the following

$\mu_E(E_H|s_H)$ = [prob(a_H) x prob(a_H sends s_H)] / [prob(a_H) x prob(a_H sends s_H) +

prob(a_L) x prob(a_L sends s_H)]

This equals

$\mu_E(E_H|s_H) = q(1)/[q(1) + (1\text{-}q)(1)] = q$

Thus, wC is a best reply to sH for the employer if

$$\mu_E(E_H|s_H)\prod(w_C|E_H) + \mu_E(E_L|s_H)\prod(w_C|E_L) \geq \mu_E(E_H|s_H)\prod(w_o|E_H) + \mu_E(E_L|s_H)\prod(w_o|E_L)$$

which equals

$$q(a_H-w_C) + (1-q)(a_L-w_C) \geq q(a_H-w_o) + (1-q)(a_L-w_o)$$

which simplifies to

$$w_o \geq w_C$$

which violates the inequality conditions stipulated, and so pooling on s_H is not an equilibrium. It is interesting to note that this result holds no matter what the probability is that a worker is a high aptitude type.

The pooling equilibrium where both types do not receive an education and the separating equilibrium where the low aptitude types receive an education while the high aptitude types do not are not relevant in both the intuitive and the analytical sense. Intuitively, these results will not ever happen, and this can be demonstrated by the same analysis done for the two relevant equilibria.

Signaling literature covers such topics as wealth serving as a signal of the ability to produce high-quality product in Okumura (2006) and retention decisions serving as a signal of worker productivity in Waldman (1990).

Regarding the issue of retention in higher education as it relates to a signaling model, there are two issues. The first is what signals identify potential students who will persist past their freshman year at Rhodes. The second is how the payoffs that Rhodes offers students can be adjusted in order to create the separating equilibrium where the students who are likely to be retained will come to Rhodes and the students who are not will not be accepted or enroll.

The payoff for the education that Rhodes offers students is a function of the combination of prestige and monetary reward that the degree confers (w_C=f(prestige); $p_H \geq p_L$) minus the cost of the Rhodes education (s_H/a_H). The first part, the prestige and monetary reward, is quantified by the average starting salary of a graduate, which is correlated to such things as the prestige of the graduate schools that other students are admitted to and attend (which possibly is also quantifiable by a salary or value-added dollar figure), and the prestige of non-profit and government jobs. The cost is the effort it takes to complete a Rhodes degree divided by the aptitude of the student, making the cost less for high-aptitude students, combined with the monetary cost of attendance (s_H=f(price of a college education, financial aid)). Therefore, there are two ways to raise the payoff for students: 1) increase the prestige of the school and thus the benefit that a Rhodes College degree confers, and 2) lower the cost of attendance.

The trick is in how to accomplish both of these things. The first part, the prestige, is determined by the payoff current graduates are already getting. That means, however, that the college will only attract those students who are in the same range of ability as those students already at the school, because these students expect to receive the same payoff based on their own assessment of their ability. There is then a conundrum; you need better students raise the payoff, but you need a higher payoff to attract the better students.

This is where part 2, lowering the cost of attendance, comes in. The payoff to a good student can be raised independently of any increase in the college's prestige by decreasing the cost of education. Obviously, making the curriculum easier and thus requiring less effort would be another way to do this, but it would come at the cost of

devaluing the product the college produces and thus eventually decreasing the prestige of the school (though there may be some evidence that the curriculum can be made easier at schools with a top reputation already, without damaging the reputation much, but this does not hold for lower reputation schools, as it is their product that is building their reputation, not their reputation sustaining their product by attracting high aptitude students). So the only viable mechanism for lowering the cost of attendance is scholarships targeted at students who are likely to be retained.

The issue then is discovering which types of students are likely to be retained, and then lowering the cost of attendance for those students to create an additional incentive for them to stay past their freshman year, and hopefully for all four years.

The proposed model for determining a student's likelihood of being retained is as follows:

Retention = f(high school GPA, SAT, high school score, gender, race, financial aid, health of the economy, home region, religion, varsity athlete, legacy, extracurricular activities)

These variables encompass Tinto's model, encompassing factors of academic and social commitment, as well as institutional and goal commitment. Different values and combinations of them serve as the signals s_H and s_L that respectively indicate the probability of being retained, put another way the ability to commit to being retained, a_H and a_L.

High school GPA, SAT, high school score, and high school extracurricular activities signal the academic commitment of a student. Gender, race, home region, religion, varsity athlete, legacy, and high school extracurricular activities signal the social

commitment of the student. Financial aid and the health of the economy alter the payoffs for a student of coming to Rhodes, and so play into the relative strength of the signals given by the other variables.

Technique

It is impossible to predict with one hundred percent certainty that an incoming student will persist from their freshman to sophomore year. Certain characteristics, however, can be identified in students before they enroll. These characteristics act as signals that will indicate whether a student has a high or low probability of persisting.

To isolate the significant pre-enrollment characteristics that act as signals of retention probability, a probit regression was run. Probit was chosen over simple OLS regression because of the binary nature of the dependent variable, and the fact that the probabilities are grouped in the high end of the distribution. Additionally, probit regression was chosen over logistic (logit) regression, the type used in the higher-education research literature on retention. This decision was made due to the tendency of logit to skew the results when there are a high number of zero outcomes in the dependent variable for any subgroup in dummy variables. Since the retention model uses many dummy variables, this problem was avoided by using probit regression.

One of the constraints of a probit regression is that there is no straightforward way to measure the goodness of fit of the model like there is for OLS regression with the R-squared statistic. Therefore, out-of-sample prediction was used to look at the success of the model at predicting retention probabilities. For reasons discussed below, this took the form of using data from first-year cohorts 2001, 2002, and 2004 to run the probit and

develop the model, and then using the data from first-year cohort 2005 as the out-of-sample segment to test the predictive power of the model.

Data

The data was assembled from the Rhodes College database using Stata. Stata is capable of pulling data directly from the database, so the code can easily be modified to replicate this study as additional years of usable data become available. This data was then manipulated and analyzed using Stata's capabilities in each of these areas.

The Rhodes database has data going back to the first-year cohort (FYC), defined as the fall that a class entered Rhodes, of 1987 (which would be the class of 1991). Records on some of the variables that were relevant were not kept until the FYC 2001, though, so the dataset that was assembled covers FYCs 2001 to 2006. Out of these years, however, two of them were unusable due to the fact the the high school GPA data was not standardized in those years, FYC 2003 and FYC 2006. Therefore, the usable data was limited to FYCs 2001, 2002, 2004, and 2005.

The summary statistics for the data actually used in the analysis are given in the following table [Fig. 3]. A binary independent variable equal to 1 if a student stayed at Rhodes from their freshman to their sophomore year begins the table. The remaining portion of the table gives the dependent variables that were chosen to measure the categories given in the theory section for the theoretical equation

Retention = f(high school GPA, SAT, high school score, gender, race, financial aid, health of the economy, home region, religion, varsity athlete, legacy, extracurricular activities)

Summary Statistics of Key Variables

	FYCs 2001, 2002, 2004, & 2005 (N = 1679)			
Variable	Mean	Std. Dev.	Range	
Retained indicator	0.87	0.33	0	1
High school GPA	3.64	0.40	1.81	4
High school score	0.80	0.21	0.04	1.00
SAT score	1267	117	950	1600
West	0.03	0.17	0	1
Midwest	0.08	0.27	0	1
East	0.08	0.27	0	1
Percent change in the Dow at time of enrollment	3.76	10.73	-7.35	20.58
Unmet need/1000	-3.76	7.83	-37.55	30.41
Black	0.05	0.21	0	1
Male	0.43	0.49	0	1
Legacy	0.10	0.30	0	1
Christian other than Presbyterian	0.52	0.50	0	1
Presbyterian	0.14	0.34	0	1
Other religion	0.16	0.37	0	1
Varsity athlete	0.27	0.45	0	1
Extracurricular activities	1.30	1.38	0	7

Fig. 3

The first is the standardized *High school GPA* that the admissions office computes from the high school transcript of students and is given on a 4.0 scale. *SAT* is the higher of either the reported SAT or the ACT score converted to SAT based on a nationally standardized conversion. *High school score* is a measure of the relative quality of all the high schools in the nation that is put together by the higher education consulting firm Human Capital and is used by college and university admissions offices across the country. It ranges from 0 to 1 with a uniform distribution. An interaction term was put in for *High school GPA* and *High school score* to weight a student's GPA by the strength of the high school. *Extracurricular activities* is simply the number of high school extracurricular activities a student listed on their college application. The signs on all of

these variables are expected to be positive, as those students who achieve more are better equipped and committed to succeed at Rhodes.

Gender is measured by the dummy variable *Male*. This is expected to be positive, as males have an incentive to stay given the significantly higher number of females on campus, an almost 2 to 1 ratio. Race is given by the dummy variable *Black* that is equal to 1 if a student identifies themselves as black, and 0 otherwise. This coefficient is expected to be positive as black students have an excellent support system once they reach campus, and also have a higher incentive to overcome information asymmetry, given Rhodes' history as a predominantly white school. Home region is measured by dummy variables for students from the *West*, *Midwest*, and *East*, with those from the South being the base group. All of these are expected to have negative signs, as students who come from further away have higher information costs in coming to Rhodes.

Religion is indicated by dummies for *Christian other than Presbyterian*, *Presbyterian*, and *Other religion*, with those who identified themselves as having no religion or being agnostic as the base group. Christians other than Presbyterian is expected to have a negative sign given the propensity for the Search and Life curriculums to make students look at the Bible in a different and potentially more adversarial way than they ever have before. Presbyterian is expected to have a positive sign, as there are lower information costs for students who grew up in the Presbyterian church and so were more likely to hear about Rhodes given its historic affiliation with the Presbyterian church. Other religion is expected to have a negative sign, as the focus of the Search and Life curriculum on the Bible gives them an increased incentive to leave.

Varsity athlete is a dummy variable that equals 1 those students who participated in varsity athletics their first year, given that they would have indicated their willingness to participate before they enroll. The sign is expected to be positive, as athletic teams provide a social network that these students are naturally integrated into. *Legacy* is a dummy variable that equals 1 for those students with a family member who went to Rhodes. The sign is expected to be positive, as those who have a family member who went to Rhodes have lower information costs as well as a greater incentive to stay.

There are two financial variables. The first, *Unmet need/1000*, measures the relative cost of attending Rhodes for a student. It is calculated by taking the computed need as determined by a standard equation the Financial Aid Office uses and then subtracting the total amount of financial aid awarded by the college. This financial aid includes scholarships, grants, loans, and workstudy from federal, state, and institutional sources. The number is then divided by one thousand to give units of thousands of dollars. A negative number for this variable means that a student was given more financial aid than they needed according to the calculation. The second financial variable is the *Percent change in the Dow at the time of enrollment*, which measures the health of the economy. It is calculated by taking the percent change in the DOW Jones Industrial Average on the last business day in April. The last business day in April was chosen because admitted students must make a decision to enroll on May 1. This measures the health of the economy and thus families's perceived ability to pay for college based on the relative health and prospects of their portfolio.

Several things stand out about these summary statistics. First, there is a wide range of variation in the data. Students's performance in the performance variables was

high at the mean, but included low performing students, as well. The overwhelming majority of students come from the South, the base group of the regional dummies. The majority identify as Christian of some denomination, though a surprising number are either not religious or agnostic, which is the base group for the religion dummies. Given the reputation of Rhodes students as very involved in extracurricular activities, it is surprising that the mean number was just over one. Perhaps, many of these students have substantial involvement in that one activity, but a more likely explanation is that there is a small number who are very involved, and many who are not involved at all. The mean amount of unmet need is negative, which is not surprising given the high number of students awarded merit-based scholarships that many times are in excess of the calculated need.

Results

To develop the model, the data from FYCs 2001, 2002, and 2004 was used to run a probit regression with the variables from the theoretic model. The results are summarized in the table below [Fig. 4]. The interpretation of the coefficients given the probit technique used is that a one-unit change in the independent variable leads to a ß-standard deviation change in the dependent variable.

The performance variables *High school GPA*, *High school score*, the *HS score with GPA interaction* term, and *SAT score* were jointly significant at the five-percent level. All except the interaction term had positive coefficient, as expected. The financial variables *Percent change in the Dow at time of enrollment* and *Unmet need/1000* were also significant with the expected signs.

The regional variables *West* and *East* are negative, as expected, but insignificant. *Midwest*, however, is negative and significant. This is most likely because Rhodes is a backup school for students from that region, as it is relatively close to home, but not overly so, as the information asymmetries are greater than for those students in the South but there is not as great an incentive to overcome them as there is for those students from further away.

Black, *Male*, and *Legacy* all had the expected sign. *Male* was not significant and small, however, suggesting that the low male-female ratio at Rhodes is not a factor in students's retention decisions. The religion dummies all had the expected sign. *Presbyterian* was the only significant one, however. This suggests that Rhodes is a welcoming place to those of any religion, but that those with a Presbyterian background in general have better information about the school.

Varsity athlete had the expected sign but was insignificant, suggesting that any differences observed in varsity athletes are non-existent when controlling for the other variables in the model. *Extracurricular activities*, on the other hand, though insignificant, had a negative sign, the opposite of what was expected. This may be that students who are overly committed in high school have the same tendency in college, but burn out.

Results

Variable	Coefficient Estimate (Std. Error)	
High school GPA	1.075	+
	(0.623)	
High school score	3.926	+
	(2.594)	
HS score with GPA interaction	-0.919	+
	(0.688)	
SAT score	0.00045	+
	(0.00046)	
West	-0.144	
	(0.274)	
Midwest	-0.486	***
	(0.161)	
East	-0.039	
	(0.183)	
Percent change in the Dow at time of enrollment	0.853	**
	(0.420)	
Unmet need/1000	-0.024	***
	(0.007)	
Black	0.548	**
	(0.275)	
Male	0.018	
	(0.102)	
Legacy	0.361	**
	(0.163)	
Christian other than Presbyterian	0.100	
	(0.132)	
Presbyterian	0.318	*
	(0.183)	
Other religion	-0.037	
	(0.162)	
Varsity athlete	0.002	
	(0.101)	
Extracurricular activities	-0.038	
	(0.038)	
constant	-3.928	*
	(2.369)	

Note: Standard errors in parentheses.
*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.
+ Joint significance at the 5 percent level.

Fig. 4

The significant variables, however, were: the performance variables *High school GPA*, *High school score*, the *HS score with GPA interaction* term, and *SAT score*; the financial variables *Percent change in the Dow at time of enrollment* and *Unmet need/1000*; *Midwest*; *Black*; *Legacy*; and *Presbyterian*. The broad lessons to be drawn

from this list is that the students who have a high probability of being retained are those that indicate they have the willingness and aptitude to work and perform well in school, the ability to pay for their education, and credible information about life both academically and socially at the school.

The model can be used to identify which groups have these characteristics. The marginal effects command in Stata may be used to get the exact unit changes for specific values in the model, but those values have to be specified, as the marginal changes will be different at any different point. The command also then gives a predicted probability of retention for a student with those characteristics. The following figure [Fig. 5] provides three examples of the predicted probabilities for students with certain characteristics. This application of the model is useful as it gives the ability to investigate specific students.

Predicted retention probabilities for some typical students

- P(HS GPA=4, SAT=1400, HS score=.3, Dow=.05, Unmet need=-5000, Southern, Black, Female, Christian, Extracurriculars=3) = **.9373**
- P(HS GPA=2.3, SAT=1000, HS score=.95, Dow=.10, Unmet need=0, East, White, Male, Christian, Athlete, Extracurriculars=5) = **.7556**
- P(HS GPA=3.0, SAT=1200, HS score=.7, Dow=-.01, Unmet need=2000, White, Midwestern, Female, Presbyterian, Legacy, Extracurriculars=2) = **.7551**

Fig. 5

Once the probit regression was run with the data from FYCs 2001, 2002, and 2004 and the model developed, the model was used to predict the probability of being retained for each student in FYC 2005. The data from FYC 2005 was then divided into quintiles based on the predicted probabilities. The actual percentages of students in each quintile who were retained were then calculated. This data is given in the table below [Fig. 6]. As is seen, the model did an accurate job of predicting those who would be

retained and those who would not. Those with the least probability of being retained left with greater frequency, while those with the highest probability of being retained stayed with greater frequency.

FYC 2005 Prediction Accuracy by Quintile

Quintile	Mean	Range		Retained	Left	% Retained
1	0.9562	0.9403	0.9854	77	6	0.9277
2	0.9252	0.9084	0.9402	79	5	0.9405
3	0.8937	0.8766	0.9084	70	13	0.8434
4	0.8514	0.8241	0.8764	73	11	0.8690
5	0.7675	0.2061	0.8240	65	19	0.7738

Fig. 6

Conclusions

The empirical model then confirms that there is a separating equilibrium that allows the reliable pre-enrollment identification of students with both high and low retention probabilities. The results teach a few things. First is that social indicators matter less than previously thought. Rather, academic factors are the most consistent predictor of retention probability. This is shown by the stability and robustness of the results throughout the evolution of the model. Next is that financial factors are key; students who have their education paid for (or to put it another way, who do not have large costs they are unable to pay for) are more likely to stay. Lastly, information costs and asymmetries matter. Those students who know more about the school coming in and so make an informed decision about whether to matriculate are more likely to stay committed to their decision.

The lessons economically from these results are that incentives and information costs matter. Students want to be academically successful because it makes their degree valuable. Therefore, they will go where they can be successful. The value of their

degree is also enhanced if it costs less to get that degree. Additionally, efforts to reduce information costs and asymmetries for prospective students are worthwhile given that those students who know more

The practical use of the model, beyond teaching us what types of students are most likely to be retained, is in identifying enrolled students with low retention probability for intervention purposes. This is especially useful given the importance and impact of academic performance. Rather than waiting for a point at which the student's academic performance may already have given reason for that student to attrit, intervention can happen during orientation, before the start of classes. Study skills and extra attention may be given to these students to help ensure their success academically.

Further research on this topic would investigate those factors that influence retention decisions post-enrollment, as well as more in-depth research on the separate components of academics, financial aid, and information costs and asymmetries.

Bibliography

Angrist, Joshua, Daniel Lang, and Philip Oreopoulos. 2006. "Lead Them to Water and Pay Them to Drink: An Experiment with Services and Incentives for College Achievement." National Bureau of Economic Research, Inc, NBER Working Papers: 12790.

Braxton, J. M. Ed. 2000. *Reworking the Student Departure Puzzle*. Nashville: Vanderbilt University Press.

Braxton, J. M., A. V. S. Sullivan, and R. M. Johnson. 1997. "Appraising Tinto's theory of college student departure." *Higher Education: Handbook of Theory and Research.* Vol. 12. Ed. J. Smart. New York: Agathon Press. 107-164).

Daula, Thomas, and Robert Moffitt. 1995. "Estimating Dynamic Models of Quit Behavior: The Case of Military Reenlistment." *Journal of Labor Economics*. 13.3: 499-522.

Dearden, Lorraine, Carl Emmerson, Chris Frayne, and Costas Meghir. 2005. "Education subsidies and school drop-out rates." Institute for Fiscal Studies, IFS Working Papers: W05/11.

Desjardins, Stephen L., Dennis A. Ahlburg, and Brian P. McCall. 2002. "Simulating the longitudinal effects of changes in financial aid on student departure from college." *Journal of Human Resources*. 37.3: 653-679.

Fullerton, Richard L. 2003. "An Empirical Assessment of US Air Force Pilot Attrition." *Defense and Peace Economics*. 14.5: 343-355.

Heath, Will Carrington. 1993. "Choosing the right pond: College choice and the quest for status." *Economics of Education Review*. 12.1: 81-88.

Hotchkiss, Julie L., Robert E. Moore, and M. Melinda Pitts. 2006. “Freshman Learning Communities, College Performance, and Retention.” *Education Economics*. 14.2: 197–210.

Huang, Ing-Chung, Hao-Chieh Lin, and Chih-Hsun Chuang. 2006. “Constructing Factors Related to Worker Retention.” *International Journal of Manpower*. 27.5: 491-508.

Johnes, Geraint and Robert McNabb. 2004. “Never Give up on the Good Times: Student Attrition in the UK.” *Oxford Bulletin of Economics and Statistics*. 66.1: 23-47.

Langbein, Laura I., and Kevin Snider. 1999. “The impact of teaching on retention: Some quantitative evidence.” *Social Science Quarterly*. 80.3: 457-472.

Levy, Stuart, and Julie Murray. 2005. “Tertiary Entrance Scores Need Not Determine Academic Success: An Analysis of Student Performance in an Equity and Access Program.” *Journal of Higher Education Policy and Management*. 27.1: 129-40.

Longden, Bernard. 2006. “An Institutional Response to Changing Student Expectations and their Impact on Retention Rates.” *Journal of Higher Education Policy and Management*. 28.2: 173–187.

Marcus, Richard D. 1989. "Freshmen Retention Rates at U.S. Private Colleges: Results from Aggregate Data." *Journal of Economic and Social Measurement*. 15.1: 37-55.

McLaughlin, G. W., Y. Perkhounkova, and J. P. Noble. 2006. “Factors related to persistence of freshmen, freshman transfers, and nonfreshman transfer students.” *AIR Professional File*. 99.

Mixon, Franklin G., Jr., and Len J. Trevino. 2005. "From Kickoff to Commencement: The Positive Role of Intercollegiate Athletics in Higher Education." *Economics of Education Review*. 24.1: 97-102.

Montmarquette, Claude, Sophie Mahseredjian, and Rachel Houle. 2001. "The Determinants of University Dropouts: A Bivariate Probability Model with Sample Selection." *Economics of Education Review*. 20.5: 475-84.

Nichols, J. L., P. M. Orehovec, and S. Ingold. 1998. "Using logistic regression to identify new 'at-risk' freshmen." *Journal of Marketing for Higher Education*. 9.1: 25-37.

Okumara, Tsunao. 2006. "Wealth as a Signal in the Search Model of Money." *International Economic Review*. 47.1: 87-106.

Robst, John, Jack Keil, and Dean Russo. 1998. "The Effect of Gender Composition of Faculty on Student Retention." *Economics of Education Review*. 17.4: 429-39.

Sacerdote, Bruce. 2000. "Peer Effects with Random Assignment: Results for Dartmouth Roommates." National Bureau of Economic Research, Inc, NBER Working Papers: 7469.

Sadler, W. E., F. L. Cohen, and L. Kockesen. 1997. "Factors affecting retention behavior: A model to predict at-risk students." *Association for Institutional Research Annual Forum*.

Singell, Larry D., Jr. 2004. "Come and Stay a While: Does Financial Aid Effect Retention Conditioned on Enrollment at a Large Public University?" *Economics of Education Review*. 23.5: 459-71.

Spence, Michael. 1973. "Job Market Signaling." *Quarterly Journal of Economics*. 87:3: 355-74.

Stock, Wendy, Aldrich Finegan, and John Siegfried. 2006. “Attrition in Economics Ph.D. Programs.” Department of Economics, Vanderbilt University, Working Papers: 0608.

Stratton, Leslie S., Dennis M. O'Toole, and James N. Wetzel. 2005. "Comparing First Term and First Year College Attrition." *Virginia Economic Journal*. 10: 29-44.

Tinto, Vincent. 1975. “Dropout from Higher Education: A Theoretical Synthesis of Recent Research.” *Review of Educational Research*. 45.1: 89-175.

Tinto, V. 1986. “Theories of student departure revisited.” *Higher Education: Handbook of Theory and Research Vol. 2*. Ed. J. Smart. New York: Agathon Press. 359-384.

Tinto, V. 1987. *Student leaving: Rethinking the causes and cures of student attrition*. Chicago: University of Chicago Press.

Tinto, V. 1988. “Stages of student departure: Reflections on the longitudinal character of student leaving.” *Journal of Higher Education*. 59: 438-455.

Tinto, V. 1993. *Leaving college: Rethinking the causes and cures of student attrition* (2nd ed.). Chicago: University of Chicago Press.

Tinto, V. 2002. “Promoting student retention: Lessons learned from the United States.” Paper presented at the 11th Annual Conference of the European Access Network, Prato, Italy: June 20.

Waldman, Michael. 1990. “Up-or-Out Contracts: A Signaling Perspective.” *Journal of Labor Economics*. 8.2: 230-50.

Wetzel, James N., Dennis O'Toole, and Steven Peterson. 1999. "Factors Affecting Student Retention Probabilities: A Case Study." *Journal of Economics and Finance*. 23.1: 45-55.

Yorke, M., B. Longden. Eds. 2004. *Retention and student success in higher education*. Buckingham: Society for Research into Higher Education and Open University Press.