

## **GRADUATION RATES AT COLLEGES AND UNIVERSITIES IN THE MIDWEST**

Jeff Anstine, North Central College  
Megan Seidman, North Central College

### **ABSTRACT**

*This research examines the relative impact of a broad variety of variables on the graduation rates of 283 four-year colleges and universities in the Midwest. It compares the relative importance of campus life (such as sports, number of student organizations and Greek life) to financial characteristics (like the percentage of students who borrow and the proportion receiving Pell Grants). It looks at these features building on previous literature that studies institutional (like type of school) and student (percent female and selectivity) characteristics. Results show that economic variables such as the socio-economic status of students is much more important in predicting graduation rates than social characteristics like the percentage of students in sororities.*

**JEL:** A20, A22, Z18

**KEYWORDS:** Graduation Rates, Colleges, Universities, Higher Education

### **INTRODUCTION**

A college degree has never been more important. High school students and their parents realize this: thus there were over twenty million students attending American colleges and universities in the fall of 2015 (National Center for Education Statistics, NCES). On average, workers with a college degree earn almost twice as much as those with only a high school degree. Thus, over a lifetime college graduates earn over a million dollars more compared to those without one (Leonhardt, 2014). In addition, individuals with a college degree report a higher sense of self-worth and satisfaction with life. Higher education in the United States is extremely important to not only all of those involved, but to the country as a whole. The increased costs over the past two decades has impacted students, potential students and their parents, with student debt now well over a trillion dollars. In addition, most companies need qualified, skilled workers with degrees that help them compete in an increasingly global world economy. There are social benefits to higher education, with it being correlated to lower crime rates and more social responsibility. Directly or indirectly related to almost all of the areas in post-secondary education is the graduation rates at colleges and universities. Policy makers want to ensure that government funds are spent wisely. Part of the Obama administration's accountability program for higher education looks at improving graduation rates. The Bill and Melinda Gates Foundation spends millions of dollars promoting college. Additionally, the basic fact that students do not attend college with the intention of flunking out or leaving before finishing. Thus, those variables that increase (or decrease) graduation rates deserves special attention. On the following pages this paper will provide a broad overview of the literature on graduation rates focusing on the articles most related to our research. We then discuss the data sources and summarize our variables. Next we explain the econometrics used and our results. We conclude the paper with how our findings add to the literature and suggestions for future research.

## LITERATURE REVIEW

There is a large and varied body of literature on graduation rates at colleges in the United States. As with most disciplines, some of the research focuses on developing mathematical models that lead to theoretical foundations explaining differences. Related literature attempts to improve models based on data. One example is Bailey (2006), who uses data mining to develop a system of predicting graduation rates at two and four year schools in Minnesota. Much of the research examines how specific programs can increase graduation rates at one particular school or a small number of institutions. Kilgo and Pascarella (2016) looked at about 4,000 students at seventeen institutions and found that undergraduate students who conducted research with professors does improve the likelihood of students completing their degrees. Unlike most anecdotal evidence in this area, they held constant other relevant variables in order to control for differences in characteristics of students. A lot has been written about the importance of mentoring, advising, new student orientation, tutoring, and other actively intensive programs for student success. White (2015) is one of many who contends that improved advising for students leads to higher levels of graduation. Lei, Gorelick, Short, Smallwood and Wright-Porter (2011) argue that cohorts (placing students in the same fields together) helps promote a support system that can improve academic outcomes.

It is well known that certain demographic groups such as first generation students, minorities, and those from lower socio-economic backgrounds not only attend college at lower rates but are also less likely to finish. Denny, Doyle, McMullin and O'Sullivan (2014) find that programs for at-risk students with financial aid improves retention better than providing aid only. Oseguera, Locks and Vega (2009) finds that specific programs aimed at helping Latinas/os overcome perceived prejudice and bias and other obstacles improves completion rates within this group. There is also debate about whether graduation rates are an accurate gauge of school quality or truly measure the value added students get from institutions. Kalsbeek and Zucker (2013) is one of the more recent studies on retention theory. They contend that market structure and schools positions are the primary driver between graduation rates. The 'brand-name' schools' that are highly selective, have high list and net prices, with top students from around the country have higher graduation rates because of the well prepared, mostly wealthy student bodies they attract.

The majority of studies analyzing graduation rates rely on case studies, anecdotal evidence, qualitative analysis or just focus on one or a small number of schools. Regression is a well-known econometric technique used for predicting particular outcomes based on certain characteristics. Thus, it is somewhat surprising that it has not been used much in the literature on forecasting retention and graduation. However, there has been some literature using regression in the area and it has grown over the past few years. Pike and Graunke (2015) use a fixed-effect model to account for variables that are not, or cannot, be included in predicting retention rates. They control for time-invariant characteristics (such as if the school is public or private), time-varying institutional (undergraduate enrollment, etcetera) and time-varying cohort (like the percentage minority and non-traditional age students) characteristics. Of the 464 schools in their data set, they find that student characteristics such as incoming standardized test scores are largely responsible for differences in retention rates.

There are very few studies that use large data sets examining graduation rates from a macro level. Anstine (2013) examined about 1,400 schools finding that the percentage of faculty that is full-time positively impact graduation rates, and that liberal arts colleges' research and masters' universities have higher rates graduation than comprehensive ones. The existence of learning communities and teaching centers at institutions of higher learning do not improve graduation rates. However, if they type of school is taken into consideration, the existence of learning communities does improve graduation rates at comprehensive universities but do not have any impact on research and masters' universities or liberal arts colleges. Most recently Horn and Lee (2016) tested the reliability and validity of regression examining almost 1,500 institutions of higher education. They have four broad categories of explanatory variables: structural (type of school, selectivity, etcetera), demographic characteristics (such as gender and ethnicity), financial (like

money spent on instruction), and contextual (degree of urbanization and unemployment rates in the surrounding area). They conclude that regression will provide reliable and valid results as long as the model is specified correctly. This research adds to the body of knowledge in a few ways. Importantly, while controlling for relevant characteristics that have been shown to impact graduation rates in previous studies, we have included different predictor variables that have not been examined before. Thus we compare the marginal impact of social variables, like the percentage of students in Fraternities, to socio-economic variables such as the affluence of the student body. In addition, we also look at how specific variables impact graduation at different schools.

## **DATA AND METHODOLOGY**

Data was compiled from about a half a dozen different sources on ‘traditional’ four year colleges and universities in the Midwest. (See the appendix for more information about classifications of colleges and geography.) From all potential schools, a significant number of schools have not provided information on percentage of faculty that is full-time, average ACT scores and other variables that contribute to differences in graduation rates. Thus, these colleges cannot be included in the analysis. Not surprisingly the majority of these schools have very low graduation rates. There are also newer schools that have just been around for a decade or two and small schools that do not report relevant information either. Thus, our analysis only examines the ‘better’ schools in the region. In addition, we have excluded: for-profit schools (for example, DeVry), those specializing in On-line degrees, specialty schools in Art (Minneapolis College of Art & Design), Business, Culinary and other trades, those admitting only upper division transfer and graduate students (Governors State University) and seminaries (Grace Bible College). Thus, the vast majority of the schools have at least some residential component, all have at least a few student organizations and all require standardized tests for admission.

Data was gathered directly from the U.S. Department of Education. The Integrated Post-Secondary Education Data System (IPEDS) gathers information from four-year institutions on graduation rates, number of student athletes and dozens of other variables. The Chronicle of Higher Education provides data on faculty salaries. We used the most recent data from the U.S News and World Report Rankings, though it should be noted that much of their data is from the U.S. Department of Education. We used it instead of IPEDS directly because it was less cumbersome to compile. In addition, some variables (such as the state) were just calculated by the authors. Initially a total of 106 variables were collected but some of them measured the same thing (percent who lived on campus and percent who lived off) so some of them were excluded. In other situations, variables were extremely correlated (levels of professors’ salaries) so only one of the variables was included (Associate Professor Salary) in our data. There were also some variables in this initial data set that we would have liked to include, such as the percentage of students that use the GI Bill to pay for college, but almost 100 schools did not report that data, so were unable to.

There was an enormous amount of correlation between many of the sports variables, such as between aid given to athletes and total expenses and revenue. Here we just kept total revenue and average salary for men’s coaches as a proxy for the importance of sporting events. Thus, there are a total of seventy-two variables in our data set. Table 1a and 1b describe the variables, provide the source of the data and give simple descriptive statistics. Note that where available, the numbers in our data set are comparable to other statistics. For example, for the 283 schools the average for the percentage of students who are female is about fifty-six percent, which is comparable to other recent studies (Anderson, 2014). Horn and Lee (2016) categorize variables in their research into four broad categories: structural (such as Carnegie classification), demographic, financial (educational expenditures), and contextual (labor market conditions). Similarly, Pike and Graunke (2015) follow a well-used format of variable classification into: time invariant institutional (such as Carnegie classification), time varying institutional (undergraduate enrollment), and time varying cohort (percentage of non-traditional age students) characteristics.

Table 1a: Description of Structural, Selectivity and Demographic Variables

Variable	Description of Variables	Data Source	Numb Obs	Min	Max	Mean	Number
GradRate	Graduation rate at each school.	US News (from IPEDS)	283	0.08	0.96	0.577	
<b>STRUCTURAL</b>							
Regional	If the school is a regional university (yes=1)	Carnegie Foundation	283	0	1	0.57	161
LibArt	If the school is a Liberal Arts College (yes=1)	Carnegie	283	0	1	0.23	64
National	If the school is a national university (yes=1)	Carnegie	283	0	1	0.20	58
Private	If the school is Private (yes=1)	Author	283	0	1	0.67	191
Urban	If the school is in an urban location (yes=1)	Author	283	0	1	0.46	130
Suburban	If the school is suburban (yes=1)	Author	283	0	1	0.25	71
Rural	If the school is in a rural location (yes=1)	Author	283	0	1	0.29	81
IA	If the school is in Iowa (yes=1)	Author	283	0	1	0.08	24
IL	If the school is in Illinois (yes=1)	Author	283	0	1	0.16	46
IN	If the school is in Indiana (yes=1)	Author	283	0	1	0.12	35
KY	If the schools is in Kentucky (yes=1)	Author	283	0	1	0.09	25
MI	If the school is in Michigan (yes=1)	Author	283	0	1	0.11	32
MN	If the schools is in Minnesota (yes=1)	Author	283	0	1	0.08	23
MO	If the school is in Missouri (yes=1)	Author	283	0	1	0.11	30
OH	If the school is in Ohio (yes=1)	Author	283	0	1	0.14	40
WI	If the school is in Wisconsin (yes=1)	Author	283	0	1	0.10	28
<b>SELECTIVITY</b>							
Retention	The percentage of first-time, full-time undergraduate students who returned to school for their second year.	USNews-IPEDS	283	0.46	0.99	0.764	
PerClsU20	Percentage of classes with fewer than 20 students	USNews-IPEDS	283	0.234	0.94	0.570	
PerClsO50	Percentage of classes with more than 50 students	USNews-IPEDS	283	0	0.62	0.041	
StudFac	Student faculty ratio	USNews-IPEDS	283	6	26	14.01	
AccpRate	Percent of students accepted out of those who applied	USNews-IPEDS	283	0.07	1	0.688	
ACT25	The ACT scores of the 25 <sup>th</sup> percentile of entering students.	USNews-IPEDS	283	14	32	20.73	
ACT75	The ACT scores of the 75 <sup>th</sup> percentile of entering students.	USNews-IPEDS	283	18	35	26.03	
Fresh10	Percentage of students who were in the top 10 percent of their high school class	USNews-IPEDS	283	0.02	0.98	0.225	
Fresh25	Percent of students who were in the top 25 percent of their high school class	USNews-IPEDS	283	0.09	1	0.474	
<b>DEMOGRAPHIC</b>							
PerFemale	Percentage of students who are female	USNews-IPEDS	283	0	1	0.557	
Over25	Percent of students older than 25 years	USNews-IPEDS	261	0	0.7	0.137	
OutState	Percent of students from another state.	USNews-IPEDS	282	0	0.93	0.274	

The first and second columns in this table lists and defines the structural, selectivity and some demographic variables. Structural variables are those that do not change over time. Selectivity variables provide an indication of the quality of each school showing variables such as average ACT scores. Demographic show student body characteristics. The third column shows the data source. The fourth column gives the number of observations, followed by the minimum and maximum values for each variable. The last two columns show the average value of each variable and the number in the category if it is a dummy variable.

To simplify the large number of variables in our data set we follow a similar format and categorized college characteristics into seven broad categories: structural, selectivity, demographic, faculty characteristics, student body characteristics, financial, and sports. In some cases, it is not perfectly clear where a variable should be categorized. For our purposes, it is not the exact area that matters, it is just classifying them to simplify because of the large number of variables. Table 1a summarizes structural, selectivity and some demographic variables. Structural variables are those that do not change over time or if they do they change slowly. The majority of the schools, fifty-seven percent, are regional. Ninety-three of the institutions are public. About half the schools are in urban locations with a quarter each in suburban and rural. The number of schools corresponds closely to state populations with the most in Illinois and least in Iowa.

Selectivity variables include the acceptance rate (most selective University of Chicago) and percentage of students in the top ten percent of their high school class (average twenty-two percent). Demographic

variables, in Table 1a and 1b, include things such as the percentage of students who are female, about fifty-six percent, which is similar to the average for all schools in the U.S. The percentage from out of state ranges from zero to ninety-three (Washington University in Saint Louis). A few were all male or all female. Table 1b provides information on variables pertaining to faculty and students, the financial viability of schools and the importance of sports. There was a lot of variability in the faculty characteristics categories. The percentage of faculty that is full-time ranged from twenty-five to one hundred percent and average salary from \$36,594 to \$117,600 (University of Chicago).

There was also a wide range in the characteristics of student bodies. The number of undergraduates go from about 500 to over 44,000 (Ohio State). The percentage of students that are part of the Greek system averages about nine percent but varies from zero to 77 percent. Financial variables illustrate the economic stability of schools (endowment measured on a per capita basis) and socio-economic status (percentage of students receiving Pell Grants). Of the students who do borrow, the average debt is about \$29,000. In order to focus on their relative importance, sports variables have been separated from other student body characteristics. The majority of schools' sports are regulated by the National Collegiate Athletic Association (NCAA) with just twenty percent adhering to National Association of Intercollegiate Athletics (NAIA) rules.

Most schools are NCAA Division III, 110, with others ranging between five and fourteen percent. Title IX is close to achieving its goal in some ways, with the average number of sports teams for women actually just slightly ahead the number for men, averaging 8.32 to 7.69. By other measures though it is not the case, such as there are more male athletes than female overall.

## RESULTS

Regressions below examine the impact different variables, such as percentage of students receiving Pell Grants, holding other influences constant, have on graduation rates at institutions of higher learning in the Midwestern U.S. Thus, the dependent variable, the graduation rate of school  $i$ , is a function of independent variables described in Tables 1a and 1b in detail and shown below in equation 1. There have already been quite a few studies examining the first four categories, so while we do look at them our focus is on the last three where there has not been much analysis using regression.

$$Graduation\ rate_i = \beta_0 + \beta_1\ Structural + \beta_2\ Selectivity + \beta_3\ Demographic + \beta_4\ Faculty + \beta_5\ Student\ Body + \beta_6\ Financial + \beta_7\ Sports + \epsilon_i \quad (1)$$

Due to multi-collinearity not all of the variables listed in Tables 1a and 1b are used in the regressions. For example, freshman in the top 10 percent of their high school class and freshman in top 25 are both highly correlated so top 25 is excluded. The number of sports for men and number of sports for women are also highly related, and therefore are combined into one variable. In addition, we used the per capita for number of athletes and organizations to control for heteroscedasticity and large variations in the numbers. Table 2 shows the first regression with all of the variables kept in their original form. Thus, they were not transformed into logs or other different mathematical forms. The majority of the variables are quantitative so we could interpret coefficients in the usual way, but we do not. Nominal variables are placed into a dummy variable format where if the characteristic exists it is identified with a one and if it does not exist it is given a zero. Hence, these coefficients are the difference between colleges and universities with the characteristic and those without it.

Table 1b: Description of Demographic, Faculty, Student Body, Financial and Sports Variables

Variable	Description of Variables	Data Source	Numb Obs	Min	Max	Mean	Number
<b>DEMOGRAPHIC</b>							
Interntl	Percentage of students from another country. Those who originated in another country	USNews-IPEDS	283	0	0.3	0.040	
Black	Percentage of students who reported Black	USNews-IPEDS	283	0	0.83	0.081	
Asian	Percentage of students who reported Asian	USNews-IPEDS	283	0	0.23	0.028	
Hispsc	Percentage of students who are Hispanic	USNews-IPEDS	283	0	0.44	0.058	
White	Percentage of students who reported White	USNews-IPEDS	283	0.25	0.97	0.727	
Other	Percentage of students who reported as Native American, Pacific Islander, Multiracial, or did not report	USNews-IPEDS	283	0	0.3	0.067	
<b>FACULTY</b>							
PerFTFac	Percent of faculty that is full time	USNews-IPEDS	283	0.249	1	0.790	
PerinstrFT	Percentage of employees on instruction, research or service who are full-time	USNews-IPEDS	283	0.109	1	0.599	
SasocProf	Average Associate Professor Salary	Chronicle of Higher Education	of 283	36549	117600	66147	
<b>STUDENT BODY</b>							
Students	Number of undergraduate students	USNews-IPEDS	283	543	44201	6004.95	
WkendCmps	Percentage of students who are on campus on the weekends	USNews-IPEDS	228	0	0.99	0.606	
PerLiveOn	Percentage of students who live in campus housing	USNews-IPEDS	278	0	1	0.502	
StudOrg	Number of student organizations	USNews-IPEDS	283	4	21	15.41	
PerFrat	Percent of male students in a Fraternity	USNews-IPEDS	268	0	0.77	0.088	
PerSor	Percent of female students in a Sorority	USNews-IPEDS	267	0	0.67	0.095	
PerStudFT	Percentage of undergraduates who attend full-time	USNews-IPEDS	283	0.007	1	0.834	
<b>FINANCIAL</b>							
PerHaveNBA	Percent determined to have financial need	USNews-IPEDS	275	0.37	1	0.717	
PerFullMet	Percent who had need fully met	USNews-IPEDS	268	0	1	0.252	
Pellgrant	Percentage of undergraduates receiving a Pell Grant	USNews-IPEDS	282	0.062	0.926	0.342	
PerBorrow	Percent of graduating students who have borrowed	USNews-IPEDS	263	0.08	0.98	0.722	
PerCapEndow	End-of-year endowment value per full-time equivalent student	USNews-IPEDS	282	507	950232	45377.85	
ALUMGvRt	Percentage of alumni who give to the school	USNews-IPEDS	282	0.01	0.51	0.128	
<b>SPORTS</b>							
NumSportsM	Number of Sports teams for Men	USNews-IPEDS	279	0	22	7.69	
NumSportsF	Number of Sports teams for Women	USNews-IPEDS	282	0	21	8.32	
MenCoachSal	Ave. salary per FTE HD Coach/ Men's	USNews-IPEDS	271	0	1408829	12119	
MenAthletes	Total Number of Athletes in Men's Sports	USNews-IPEDS	272	0	941	281.27	
NCAADivIA	The school is in Division I – A of the NCAA	USNews-IPEDS	283	0	1	0.11	30
NCAADivIAA	The school is in Division I – AA	USNews-IPEDS	283	0	1	0.05	15
NCAADivIAAA	The school is in Division I – AAA	USNews-IPEDS	283	0	1	0.07	19
NCAADivII	The school is in Division II of the NCAA	USNews-IPEDS	283	0	1	0.17	48
NCAADivIII	The school is in Division III of the NCAA	USNews-IPEDS	283	0	1	0.39	109
NAIADivI	The school is in Division I of the NAIA	USNews-IPEDS	283	0	1	0.06	18
NAIADivII	The school is in Division II of the NAIA	USNews-IPEDS	283	0	1	0.14	40

The first and second columns in this table lists and defines demographic, faculty, student body, financial and sports variables. Faculty variables describe the importance of faculty to the schools. Student body variables are those that indicate the connectedness of students to their campus. Financial variables show the socio-economic status of students and sports shows the level of participation. The third and fourth columns show the data source, number of observations, followed by the minimum and maximum values for each variable. The last two columns show the average value of each variable and the number in the category if it is a dummy variable.

If there are more than two options the excluded category is listed (regional is excluded for type of institution, etc.). A total of forty-eight variables were included in the regression. The variables we chose were selected based on findings in previous research, but primarily came from studies with anecdotal evidence, not

necessarily using data and theory in existing literature. Regressions by Horn and Lee (2016) show that the type of school, gender and other variables are important in predicting graduation rates, so we include these. Other research that does not use regression analysis has suggested that the importance of sports to students may influence college success. There is also lots of evidence that socioeconomic variables are extremely important.

Table 2: Regression no Variables Transformed: Dependent Variable: Graduation Rate

Independent Vars	Coefficients	Stand Error	T-Statistics
Intercept	0.045	0.137	0.331
LibArt <sup>a</sup>	0.034	0.016	2.057**
National	0.004	0.014	.291
Private (yes =1)	0.108	0.021	5.141***
Urban <sup>b</sup>	-0.007	0.011	-0.636
Suburban	-0.003	0.013	-0.192
IA <sup>c</sup>	0.023	0.021	1.102
IN	-0.013	0.021	-0.654
KY	-0.038	0.022	-1.700*
MI	-0.023	0.018	-1.302
MN	0.005	0.023	0.234
MO	0.008	0.021	0.370
OH	-0.009	0.017	-0.528
WI	0.000	0.021	-0.020
PerClsU20	-0.034	0.045	-0.755
AcceptRt	-0.050	0.033	-1.510*
ACT25	0.013	0.004	3.236***
Fresh10	0.099	0.077	1.287
PerFemale	0.159	0.067	2.367**
Over25	-0.102	0.059	-1.720**
OutState	-0.046	0.031	-1.477
Interntl	0.067	0.127	0.528
Black <sup>d</sup>	-0.203	0.093	-2.180**
Asian	-0.120	0.195	-0.615
Hispsc	0.112	0.116	0.966
Other	-0.015	0.108	-0.140
PerFTFac	0.087	0.057	1.514
SalAssocProf	-2.159E-7	0.000	-0.305
Students	3.107E-6	0.000	2.474***
PerLiveOn	0.061	0.033	1.847**
StudOrgPCap	-2.406	1.864	-1.291
PerFrat	-0.013	0.086	-0.148
PerSor	0.020	0.081	0.253
PerStudFT	0.001	0.001	1.920**
PerHaveNBA	-0.143	0.074	-1.936**
PellGrant	-0.001	0.001	-2.411***
PerBorrow	0.132	0.052	2.524***
PerCapEndow	-6.615E-9	0.000	-0.104
AlmnGivRt	0.259	0.099	2.619***
MenCoachSal	8.891E-9	0.000	0.274
NCAADivIA <sup>e</sup>	-0.014	0.025	-0.578
NCAADIVIAA	0.007	0.020	0.361
NCAADIVIAAA	-0.017	0.019	-0.889
NCAADivII	0.014	0.014	0.951
NAIADivI	-0.030	0.025	-1.178
NAIADivII	-0.008	0.015	-0.490
NumSports	0.004	0.001	2.883***
MenAthPerCap	-0.462	0.133	-3.471***
WomenAthPCap	0.220	0.240	0.914
Number of Obs: 283	R-squared: 0.899	F-statistic: 31.28	

*This table shows the regression with the graduation rate as the dependent variable as a function of many explanatory variables with all of the variables in their original form. For the qualitative variables with more than two options the excluded categories are as follows: a Regional, b: Rural, c: Illinois, d: White, e: Division III. The levels of statistical significance are: \* at the 10% level, \*\* at the 5% level, \*\*\* at the 1% level.*

We included new variables for regressions that best follow earlier work in the area but have not been analyzed using econometrics in this manner. Thus, the regressions control for those variables that are known to impact graduation rates while adding new unexamined factors in the regression. We also ran a step-wise regression that suggested including most of the variables that we used. Since there is no theoretical rational to the variables chosen with this method, we are using regressions based on existing knowledge. The R squared for the simple Ordinary Least Squares (OLS) regression is 0.899, showing that practically ninety percent of the variance in graduation rates is explained by our variables. R squared in

previous literature tended to range around 0.65, with a minimum of 0.44 (Gansmer-Topf and Schuh, 2006) and a maximum of 0.85 (Scott, et al., 2006). Given that our data has more and new variables than earlier research, it is not surprising that we can explain a large amount of school's graduation rates. The big F-statistic, 31.35, is not surprising either.

The majority of the control variables have the expected sign and are statistically significant. Most of the structural variables (type of school, etc) are similar to results in existing literature. There are small differences in graduation rates between some states, with Kentucky having lower graduation rates than Illinois, but overall the differences are not statistically significant. The results of the regression for selectivity control variables is also consistent with other studies. Standardized test scores are a huge determinant of a school's graduation rate. Results of the demographic and faculty variable categories are also consistent with previous literature. Schools with a higher percentage of black students have lower graduation rates. Institutions with a higher percentage of female students graduate at higher rates. And schools with a larger percentage of non-traditional students (as proxied by the percentage of students over twenty-five) also have less academic success.

The focus of this research is on the relative importance of factors that have not been in previous regression analyses: student body, financial and sports categories. Previous anecdotal studies have shown that the more engaged students are with college, the more likely they are to graduate (Denny, et al., 2014). Controlling for other variables, our results confirm this. The percent of students that live on campus and are full-time students demonstrates engagement. Both of these variables are positive and statistically significant. The alumni giving rate is positive and statistically significant. It is likely that schools with higher giving rates had students that were engaged and thus give back at higher rates than those who were less so. The importance of social characteristics showing how involved students are in college depends on the type of activity. The number of student organizations (on a per capita basis), the percentage of students in fraternities and sororities are not statistically significant at any level. Thus, at least for Midwestern institutions, the importance of these is not relevant for schools' graduation rates.

One might expect that bigger, 'name' schools that are recognized by their Division I sports programs would have higher graduation rates than other institutions. Our results show that this is not the case in the Midwest with there being no statistically significant difference in the graduation rates of schools across all divisions. In addition, holding constant the level of play, the higher the number of sports, the higher the graduation rate. It is likely that the student-athletes at all levels are engaged, thus making no difference. The regression in Table 2 confirms other studies showing that schools that enroll higher numbers of disadvantaged students graduate at lower levels than institutions that have more students from well to do families.  *Holding constant all other influences*, colleges that have a higher percent of students with need based aid and those receiving Pell Grants have lower graduation rates. Unfortunately, no matter what type of school, the demographic characteristics, etcetera, schools that have more students from lower socio-economic backgrounds are less successful in having their students' complete college.

Some variables were found to be non-linear. For example, a simple graph examining the percentage of freshman that are in the top 10 percent of their high school class appeared to be increasing at a decreasing rate. Thus, we squared it and also included that. Regressions showed the untransformed variable to be positive and the squared on to be negative, both statistically significant, confirming this. Dozens of other regressions were estimated with some variables transformed into logarithmic and other forms. After comparing adjusted R squared with these, while continuing to base regressions on existing knowledge and theory, we found the best regression. Table 3 shows a regression with six variables in logarithmic form: faculty salary, number of students, percent living on campus, percent who borrowed, per capita endowment and percent of alumni giving, with the others not transformed. Results are very similar to those in Table 2. R squared is 0.896, compared to 0.899, again illustrating that a large percentage of the variation in graduation rates in the Midwest is explained by our variables. The adjusted R squared is almost identical



too and both have very similar F statistics. The sign and level of statistical significance on the majority of the variables in Table 3 are similar to those in Table 2 illustrating that our results are robust across different specifications. While there were some small changes in some coefficients, most remained very close, for example the sign on the Kentucky variable went from -0.038 to -0.039 and the level of significance stayed at ten percent. Thus, the results on the structural, such as private schools having a higher graduation rate than public and demographic, such as schools with a higher percentage of females having a higher graduation rate, remain consistent with earlier research.

Table 3: Regression with Some Variables Transformed: Dependent Variable: Graduation Rate

Independent Vars	Coefficients	Stand Error	T-statistics
Intercept	0.142	0.509	0.280
LibArt <sup>a</sup>	0.013	0.017	0.785
National	-0.011	0.014	-0.751
Private (yes=1)	0.084	0.022	3.856***
Urban <sup>b</sup>	0.004	0.011	0.320
Suburban	0.009	0.013	0.674
IA <sup>c</sup>	0.029	0.021	1.393
IN	-0.001	0.021	-0.059
KY	-0.039	0.022	-1.769*
MI	-0.018	0.018	-0.976
MN	0.024	0.023	1.043
MO	0.001	0.021	0.048
OH	0.001	0.017	0.055
WI	0.022	0.021	1.070
PerClsU20	0.005	0.049	0.098
AcceptRt	-0.082	0.036	-2.248**
ACT25	0.011	0.004	2.934***
Fresh10	0.320	0.119	2.680***
Fr10Sqr	-0.259	0.117	-2.214**
PerFemale	0.147	0.068	2.182**
Over25	-0.046	0.059	-0.775
OutState	-0.055	0.030	-1.825*
Interntl	-0.005	0.134	-0.040
Black <sup>d</sup>	-0.196	0.098	-1.996**
Asian	-0.187	0.193	-0.971
Hispsc	0.409	0.133	3.080***
Other	-0.011	0.110	-0.103
PerFTFac	0.066	0.057	1.160
LnSalary	-0.009	0.049	-0.193
LnStudents	0.039	0.016	2.497**
StudOrgPCap	-1.338	2.082	-0.643
LnLivOnCamp	0.025	0.013	1.964**
PerFrat	0.027	0.083	0.323
PerSor	-0.042	0.079	-0.535
PerStudFT	0.001	0.001	0.899
PerHaveNBA	-0.184	0.073	-2.516**
PellGrant	-0.001	0.001	-2.136**
LnBorrow	0.012	0.026	0.479
LnEndow	0.007	0.007	0.999
LnGiving	0.033	0.013	2.578**
MenCoachSal	1.776E-8	0.000	0.549
NCAADivIA <sup>e</sup>	-0.020	0.024	-0.854
NCAADIVIAA	-0.012	0.020	-0.599
NCAADIVIAAA	-0.043	0.018	-2.409**
NCAADivII	-0.006	0.014	-0.428
NAIADivI	-0.023	0.025	-0.904
NAIADivII	-0.018	0.015	-1.172
MenAthPerCap	-0.320	0.134	-2.394**
WomenAthPCap	0.492	0.235	2.097**
Number of Obs: 283	R-Squared: 0.896	F-statistic: 29.14	

*This table shows a regression with the graduation rate as the dependent variable as a function of independent variables with some transformed into logs. For the qualitative variables with more than two options the excluded categories are as follows: a Regional, b: Rural, c: Illinois, d: White, e: Division III. The levels of statistical significance are: \* at the 10% level, \*\* at the 5% level, \*\*\* at the 1% level.*

The major exception to the results remaining similar is the coefficient for the percentage of students that borrow. There is quite a bit of literature on the effects of borrowing to finance a college education but it excludes the impact that other variables may have. So our research sheds a little light on this, by controlling for student body and other characteristics. The coefficient is positive and significant at the five percent level when it is not transformed but is not statistically significant at all when it is logged. Dwyer et. al.

(2012) find on a micro level that there is an optimal level of debt and the relationship between it and graduation is not linear. Though much more needs to be analyzed, our results confirm this on a macro level. In addition to the similar results with our control variables, the student body, financial, and sports variables that is the focus of this paper also mostly stay the same. Holding constant other variables, the percentage of students in fraternities and sororities does not increase, or decrease, graduation rates. On a per capita basis, schools with more female athletes (holding constant the level of play) have a higher graduation rate. Again, socio-economic variables are highly important in determining graduation rates at colleges and universities in the Midwest. The higher the percentage of students that have need based aid, the lower the graduation rates. In our data, this ranged from a low of thirty-seven percent (Miami University, with the University of Michigan close at thirty eight percent) to a high of one hundred percent at a few schools. The average for all the schools is seventy-two percent with most of the schools within fifteen points of this. The average percentage of students receiving Pell Grants is almost exactly half that of students getting need based aid at thirty-four percent. The dispersion of the data is higher (as measured by standard deviations) with a range from six percent (Washington University in Saint Louis) to ninety-three percent (Drury University). At those schools with a higher percentage of students getting Pell Grants graduation rates are lower.

## CONCLUSION

The goal of this paper was to compare the relative impact on graduation rates at colleges and universities of variables that have not been examined, holding constant dozens of other factors that have already been shown to influence them. We gathered data from different sources on about three hundred schools then used regression analysis to isolate how important each variable was. Institutions that have a higher percentage of students living on campus and likely feel more of a part of the college have higher graduation rates than schools with lower percentages. The socio-economic status of students in colleges plays a huge role in determining outcomes. Using regression our results confirm that institutions that enroll higher percentages of students from disadvantaged households have lower graduation rates. While this research uses dozens of variables, thus allowing us to control for many differences in schools, there are other variables that are not in our data set. For example, specific measures of how engaged students are with professors is difficult to quantify and is not included. In addition, our focus is on quantitative research and excludes qualitative factors. Future research should look at examining other variables that we were not able to obtain and combine qualitative research with the use of econometrics.

## APPENDIX

We are examining colleges and universities, that is, post-secondary institutions. The Integrated Post-Secondary Education Data System (IPEDS) defines a postsecondary institution as an organization that is open to the public and has as its primary mission the provision of post-secondary education or training beyond the high school level, (NCES, 2016). Colleges and universities provide post-secondary education and while often used interchangeably are different. Officially a college is an educational establishment for higher or professional education. A college is also an independent part of a university. A university is an educational institution, composed of one or more colleges and graduate schools that provides instruction and facilities for research in many branches of advanced learning and awards degrees. While we recognize the difference we use the words interchangeably. There are slight differences in definitions of the Midwest. The Encyclopedia Britannica calls the “Middle West, also called Midwest, or North Central States, region, northern and central United States, lying midway between the Appalachian and Rocky Mountains and north of the Ohio River and the 37th parallel. The Middle West, as defined by the federal government, comprises the states of Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota and Wisconsin.” Kansas, Nebraska, North Dakota, and South Dakota are also considered part of the Great Plains and are not included but due to its proximity to the other states we have included Kentucky instead.

## REFERENCES

Anderson, Nick, *Washington Post*, March 26, 2014

Anstine, Jeff. (2013) "Graduation Rates at U.S. Colleges and Universities: A Large Data Set Analysis," *Business Education & Accreditation* vol. 5(2) p. 55-64

Bailey, Brenda L. (2006) "Let The Data Talk: Developing Models To Explain IPEDS Graduation Rates." *New Directions for Institutional Research* vol. 131, p. 101-115

Denny, Kevin, Orla Doyle, Patricia McMullin and Vincent O'Sullivan. (2014) "Money, Mentoring And Making Friends: The Impact Of A Multidimensional Access Program On Student Performance," *Economics of Education Review* vol. 40 p. 167-182.

Dwyer, R. E., McCloud, L., & Hodson, R. (2012) "Debt and Graduation from American Universities," *Social Forces*, vol. 90(4) p. 1133-1155.

Gansmer-Topf, A. M., and Schuh, J. H. (2006) "Institutional Selectivity and Institutional Expenditures: Examining Organizational Factors that Contribute to Retention and Graduation" *Research in Higher Education* vol. 47 p. 613-642

Horn, Aaron, and Giljae Lee. (2016) "The Reliability and Validity of Using Regression Residuals to Measure Institutional Effectiveness in Promoting Degree Completion," *Research In Higher Education* vol. 57(4) p. 469-496

Kalsbeek, David H., and Brian Zucker. (2013) "Reframing Retention Strategy: A Focus on Profile." *New Directions for Higher Education* vol. 161 p. 15-25

Kilgo, Cindy, and Ernest Pascarella. (2016) "Does Independent Research with a Faculty Member Enhance Four-Year Graduation and Graduate/Professional Degree Plans? Convergent Results with Different Analytical Methods." *Higher Education* vol. 71(4) p. 575-592

Lei, Simon, Denise Gorelick, Keily Short, Lauren Smallwood and Karen Wright-Porter. (2011) "Academic Cohorts: Benefits And Drawbacks Of Being A Member Of A Community Of Learners." *Education* vol. 131(3) p. 497-504

Leonhardt, David, *New York Times*, May 27, 2014

National Center for Education Statistics, 2016, retrieved from <http://nces.ed.gov/ipeds/>

Oseguera, Leticia, Angela M. Locks, and Irene I. Vega. (2009) "Increasing Latina/O Students' Baccalaureate Attainment: A Focus On Retention," *Journal of Hispanic Higher Education* vol. 8(1) p. 23-53

Pike, Gary, and Steven Graunke. (2015) "Examining the Effects of Institutional and Cohort Characteristics on Retention Rates," *Research In Higher Education* vol. 56(2) p. 146-165

Scott, M., Bailey, T., Kienzl, G. (2006) "Relative Success? Determinants of College Graduation Rates in Public and Private Colleges in the U.S." *Research In Higher Education* vol. 47 p. 249-279

U.S. News & World Report, retrieved from <http://colleges.usnews.rankingsandreviews.com/best-colleges>

White, Eric R. (2015) "Academic Advising In Higher Education." *JGE: The Journal of General Education* vol. 64(4) p. 263-277

### **ACKNOWLEDGEMENTS**

We would like to thank North Central College which provided a summer grant for Faculty-Student scholarship that supported part of this research. We are also grateful to the two anonymous reviewers and the editor for helpful comments that improved this paper. Any remaining errors are the authors'.

### **BIOGRAPHIES**

Jeff Anstine is a Professor of Management. He teaches a wide range of classes from econometrics to international business. His research examines topics in higher education and on environmental issues. He can be reached at North Central College in Naperville IL, 60540, [jdanstine@noctrl.edu](mailto:jdanstine@noctrl.edu), 630-637-5476.

Megan Seidman is a student at North Central College, majoring in Economics, Psychology and International Business with a Spanish minor. She can be reached at [mseidman@noctrl.edu](mailto:mseidman@noctrl.edu).