

WHY UNIVERSITIES HAVE DIFFERENT GRADUATION RATES THAN PREDICTED: A RESIDUAL ANALYSIS APPROACH

Jeff Anstine, North Central College

ABSTRACT

This paper examines differences in graduation rates at colleges and universities using both quantitative and qualitative analysis. Building on previous research this paper first uses regression equations to predict what variables influence graduation rates. After controlling for dozens of variables such as standardized test scores, this analysis then looks directly at those schools with significantly higher or lower predicted values. These schools are examined to see what variables might be important in determining graduation rates that are not easily quantified or are just not in data sets. In the quantitative section, results confirm that socioeconomic status, selectivity and other well-established variables highly influence graduation rates. The qualitative research suggests that information such as how well known schools are may influence outcomes, such as those with visible sports teams over-perform smaller, less well-known schools that do not have a visibility beyond academics.

JEL: A20, A22, Z18

KEYWORDS: Graduation Rates, Colleges, Universities, Higher Education

INTRODUCTION

As tuition rates at institutions of higher education in the United States have soared over the past two decades, students, parents and the general public have increasingly questioned the value of a college degree. Like it or not, stakeholders in academia need to address the legitimate concerns about affordability, retention, student debt, job opportunities and other challenges facing colleges and universities in the United States. There are a variety of approaches that can be used to examine these issues and potential solutions for them. Various disciplines use different techniques to study the organizations, including those in higher education.

The discipline of management examines all organizations, corporations, government agencies and non-profit organizations. It broadly looks at effectiveness, picking and obtaining organizational goals and efficiency, how well resources are used. In higher education, efficiency can be calculated with items like student to faculty ratios, teaching load, dollars spent per student credit hour and other numerical measures. Recently, in the *Chronicle of Higher Education*, Tugend (2019) discusses universities in the U.S. that are most efficient. In addition, while it is far from a perfect metric, effectiveness can be evaluated by graduation rates at schools after accounting for structural, selectivity and other institutional variables that vary significantly across institutions. (Comparing an Ivy League school with enormous endowments and other means to Historically Black Colleges and Universities with far less resources is like comparing apples and oranges.)

This research is unique and contributes to the existing literature because it combines elements of both quantitative and qualitative analysis whereas all previous investigations only use one. Prior quantitative

analysis uses regression to explain which variables impact graduation rates but do not delve into what missing variables in their models may cause differences. Qualitative research only focuses on observable characteristics at one or a few schools. This inquiry is the first to combine the two by first running regressions to determine variables that influence graduation rates and then looking at residuals of specific colleges and universities to see what variables not in the model explain the high or low difference from predicted graduation rates. While this research is unique, it is the third in a series. In the first paper I had a national data set with about 1,400 schools. This paper uses the same data set, but differently from the second paper in this series that is described in the data section.

In the following pages, I provide an overview of information on the value of college, then provide a literature review for both quantitative and qualitative research. I describe the data used in the analysis, followed by the results of the regressions. Next, I use the information on the schools that have the largest residuals to examine potential variables that these schools have, such as having four credit-hour classes, which are not in the model that might explain differences in their predicted and actual graduation rates. I finish the paper with how my results add to the literature and how qualitative analysis might add to the quantitative research in the future.

LITERATURE REVIEW

Over the past few decades there has been quite a bit of research looking at factors influencing colleges and universities graduation rates. Recently, there is a growing body of information examining Returns on Investment (ROI) and what colleges offer the best value for students. Over the past few years, Money Magazine (Mulhere and Glum, 2018) has ranked schools according to the best colleges for your money. They rate schools based on three categories: educational quality, affordability and alumni financial success. Affordability is obviously direct expenses, alumni success is the monetary premium the schools graduates have over other similar schools and quality is based on graduations rates compared to other comparable schools. Thus, schools that cost less and have higher graduation rates compared to other similar schools with comparable demographics are better values than their peers are.

In an attempt to make sure that funding is used well, some states are tying money given to colleges based on the schools graduation rates. Hester and Ishitani (2018) try to address the effectiveness of this by looking at different expenditures, on instruction, research, public service, academic support and student services. Using data from the Integrated Postsecondary Education Data System (IPEDS) and then controlling for variables like standardized test scores and region, they found that only spending on instruction increased graduation rates at public colleges and universities. There was no significant increase in graduation rates with more money spent on research, public service or surprisingly money spent on student services.

This paper uses both quantitative analysis with regression and qualitative investigation by looking directly at specific colleges campuses and their websites. Thus, the next part of the literature review examines ‘hard’ research and the last section will provide an overview of some of the ‘soft’ papers. It is interesting, and not surprising, that the quantitative research is by physicists’, mathematicians’ and economists’ while the qualitative investigations is by people in sociology and psychology.

Anstine (2013) studied approximately 1,400 colleges and universities and found that the percentage of faculty that is full-time positively affect graduation rates, and that liberal arts colleges’ research and masters’ universities have higher rates graduation than comprehensive ones. Learning communities and teaching centers at schools do not improve graduation rates. However, if they type of institution is broken down, the existence of learning communities does improve graduation rates at comprehensive universities but not at other types of schools.

Pike and Graunke (2015) provide one of the few studies on factors impacting graduation rates using sophisticated econometrics. One problem with using regression analysis is that there are likely important explanatory variables that could or should be included in the models but are unavailable. The authors do control for variables that have been shown to be extremely important such as standardized test scores and if the school is public or private. Then they use a fixed effect model to control for other factors that are not included.

Ober et. al. (2018) also use IPEDS data but narrow the schools to some two-year and approximately 500 four-year public institutions only from the years 2000 to 2015. They use a logistic growth function model that has been shown to accurately predict graduation rates at colleges and universities. In addition to confirming that family variables such as socioeconomic status and ethnicity highly impact graduation rates, schools with higher percentages of full-time students also have higher graduation rates than those with lower percentages of full-time students.

One of the most comprehensive studies on factors predicting graduation rates was by Horn and Lee (2016). They looked at approximately 1,500 colleges and universities. They broke the explanatory variables into four categories; structural (characteristics that do not change over time), demographic, financial and contextual (characteristics of the community around the school such as degree of urbanization and unemployment rates). Their results confirm that private schools have higher graduation rates than public and the more selective a school is, the higher its graduation rate.

There are many more papers looking at graduation rates at institutions of higher education using qualitative than quantitative analysis. Although many of these papers use data, it is very different from the research with large data sets. One example is Scott (2017) who asked twenty-one psychology students in upper level seminar classes about their perceptions of a college degree. The students who were predominantly white and economically well-off said that they were confident that their education prepared them for future work.

Einbinder (2018) examines eight faculty members at California State University Dominguez Hills after a one-semester experience in one type of High Impact Practice (HIP). The faculty, including the author, were put into a Faculty Learning Community where they were trained to incorporate collaborative teaching techniques into their courses and then say if it helped students learn more than they would otherwise. Seven of the faculty said they were able to incorporate the technique in their classes and four of them thought that it made them better teachers, while the other three did not think it improved their teaching.

Both Scott (2017) and Einbinder (2018) discuss the limitations of qualitative research. One is that the data uses convenience samples, so the students and/or faculty involved in the studies are not representative of not just the population in general but not even of the schools where the study is done. In addition, the studies tend to have very small sample sizes so projecting the results to other students and faculty is not applicable. Thus, assessing the validity and reliability of the studies is difficult, if not impossible.

The research in this paper adds to the literature by first conducting quantitative analysis that is valid and reliable. First, I estimate regressions controlling for relevant characteristics that have been shown to impact graduation rates in previous studies, like socio-economic status, standardized test scores, type of school and ethnicity. Then using these variables, I compare each schools actual graduation rate to its predicted one. Then, I look at the schools with the largest differences, both higher and lower and conduct qualitative analysis, looking directly at these schools to find characteristics of importance that are not in the data used for the regression analysis.

DATA AND METHODOLOGY

Data for this paper was gathered and used based on previous research in this area. It is from the same source as Anstine and Seidman (2017) but is slightly different. Schools Average ACT is in this data set while the ACT scores of students in the 25th and 75th percentiles was used in the earlier paper. Thus, based on the information in the literature review, I included variables that have the largest impacts on graduation rates while also including others that matter to control for the largest factors. Data was gathered from a variety of sources starting with the U.S. Department of Education. All data was gathered in 2016. The Integrated Post-Secondary Education Data System (IPEDS) gathers information from four-year institutions on graduation rates, retention, standardized test scores and dozens of other variables.

In addition, some variables were calculated directly by the author, such as if the school is public or private and the location of each school, urban, suburban or rural. All of the information was compiled in 2016. Since gathering data from all of the colleges and universities in the United States is a cumbersome, time-consuming process, the author narrowed the data to institutions of higher education in the Mid-West only. Thus, the author identified all four-year schools in Iowa, Illinois, Indiana, Kentucky, Michigan, Minnesota, Missouri, Ohio and Wisconsin. Also, The Chronicle of Higher Education provided data on faculty salaries. The data set ended up with 283 schools, thus 283 observations. Specialty schools (Culinary, Art and Design), Seminaries, for-profit institutions, and other non-traditional schools were excluded from the analysis.

I started with over 100 variables. Due to multicollinearity and overlap of information not all were included in the analysis. Variables such as the percentage of students that use the GI Bill were excluded because many schools did not report this. Due to the large number of variables, the descriptive statistics are broken into two tables. Table 1 summarizes time invariant variables and those related to quality. Time invariant are variables that do not change over time such as the type of school, or might change but would do so very slowly such as location measured by population density. Most schools, one hundred sixty-one are regional. Over two-thirds, sixty-seven percent are private. Approximately twenty-five percent of the schools are in rural areas and suburbia with just under fifty percent in cities. The number of schools corresponds closely to state populations with the most schools in Illinois and the fewest in Iowa.

Table 2 provides more information on different variables. Demographic characteristics provide information on the percentage of students in each school who are female, over twenty-five, out of state or from a country other than the U.S and ethnicity. Faculty information shows the percentage of faculty and employees who are full time and the average salary for Associate Professors. The student variables are intended to capture how ‘connected’ students are to their campus, such as the percentage that are on campus on the weekends, percent who live on campus and percent full-time. The financial variables provide an indication of how well-off the student body is and how stable the schools finances are. The percentage of students who receive Pell Grants shows the socio-economic status of the student body and Per Capita Endowment can be a proxy for the wealth of a school.

Table 1: Description of Variables

Variable	Description of Variables	Data Source	Numb Obs	Min	Max	Mean	Number
GradRate	Graduation rate at each school.	IPEDS	283	0.08	0.96	0.577	
TIME INVARIANT							
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
QUALITY							
Retention	The percentage of first-time, full-time undergraduate students who returned to school for their second year.	IPEDS	283	0.46	0.99	0.764	
PerClsU20	Percentage of classes with fewer than 20 students	IPEDS	283	0.234	0.94	0.570	
PerClsO50	Percentage of classes with more than 50 students	IPEDS	283	0	0.62	0.041	
StudFac	Student faculty ratio	IPEDS	283	6	26	14.01	
AccpRate	Percent of students accepted out of those who applied	IPEDS	283	0.07	1	0.688	
AveACT	The Average ACT scores of entering students.	IPEDS	283	16	33	23.38	
Fresh10	Percentage of students who were in the top 10 percent of their high school class	IPEDS	283	0.02	0.98	0.225	
Fresh25	Percent of students who were in the top 25 percent of their high school class	IPEDS	283	0.09	1	0.474	

The first and second columns in this table lists and defines the time invariant and quality variables. Time invariant variables are those that do not change over time. Quality variables provide information on retention, size of classes and average ACT scores. Data source is listed next, followed by the lowest and highest values. The last two columns average values and the number of observations if it is a categorical variable. This data is similar to that used in earlier research in a paper by Anstine and Seidman (2017).

Table 2: Description of Variables

Variable	Description of Variables	Data Source	Numb Obs	Min	Max	Mean
DEMOGRAPHIC						
PerFemale	Percentage of students who are Female	IPEDS	283	0	1	0.557
Over25	Percentage of students who are over 25	IPEDS	283	0	0.7	0.137
OutState	Percentage of students who are out of state	IPEDS	283	0	0.93	0.274
Interntl	Percentage of students from another country. Those who originated in another country	IPEDS	283	0	0.3	0.040
Black	Percentage of students who reported Black	IPEDS	283	0	0.83	0.081
Asian	Percentage of students who reported Asian	IPEDS	283	0	0.23	0.028
Hispsc	Percentage of students who are Hispanic	IPEDS	283	0	0.44	0.058
White	Percentage of students who reported White	IPEDS	283	0.25	0.97	0.727
Other	Percentage of students who reported as Native American, Pacific Islander, Multiracial, or did not report	IPEDS	283	0	0.3	0.067
FACULTY						
PerFTFac	Percent of faculty that is full time	IPEDS	283	0.249	1	0.790
PerinstrFT	Percentage of employees on instruction, research or service who are full-time	IPEDS	283	0.109	1	0.599
SasocProf	Average Associate Professor Salary	Chronicle of Higher Education	of 283	36,549	117,600	66,147
STUDENT						
Students	Number of undergraduate students	IPEDS	283	543	44,201	6,005
WkendCmps	Percentage of students who are on campus on the weekends	IPEDS	228	0	0.99	0.606
PerLiveOn	Percentage of students who live in campus housing	IPEDS	278	0	1	0.502
PerStudFT	Percentage of undergraduates who attend full-time	IPEDS	283	0.007	1	0.834
FINANCIAL						
PerHaveNBA	Percent determined to have financial need	IPEDS	275	0.37	1	0.717
PerFullMet	Percent who had need fully met	IPEDS	268	0	1	0.252
Pellgrant	Percentage of undergraduates receiving a Pell Grant	IPEDS	282	0.062	0.926	0.342
PerBorrow	Percent of graduating students who have borrowed	IPEDS	263	0.08	0.98	0.722
PerCapEndow	End-of-year endowment value per full-time equivalent student	IPEDS	282	507	950,232	45,378
ALUMGvRt	Percentage of alumni who give to the school	IPEDS	282	0.01	0.51	0.128

The first and second columns in this table lists and defines demographic, faculty, student and financial variables. Demographic variables show student body characteristics. Faculty variables provide an overview of faculty. Student variables are those that indicate how 'connected' students are to their school. Financial variables provide an indication of how well off students are. The data source and number of observations are in the third and fourth columns. The last three columns provide the smallest, largest and average values. This data is similar to that used in earlier research in a paper by Anstine and Seidman (2017).

RESULTS

Quantitative Analysis

The regressions in Table 3 and Table 4 show the impact that different variables, such as Average Act scores have on graduation rates. The graduation rate of school i , depends on variables listed in Tables 1 and 2 and is presented in equation 1.

$$Graduation\ rate_i = \beta_0 + \beta_1 Time\ Invariant + \beta_2 Quality + \beta_3 Demographic + \beta_4 Faculty + \beta_5 Students + \beta_6 Financial + \epsilon_i \tag{1}$$

There is a high degree of collinearity between some variables, thus only freshman in the top 10 percent of class was included. The percentage of students having need based aid and the percentage of students getting

Pell Grants are also highly correlated, thus only Pell Grants is included. A total of thirty-one variables were ultimately included in the regression. The variables I picked were selected based on findings in previous research but also include new variables such as Per Capita Endowment.

The regression in Table 3 is consistent with all other literature in the area with respect to important variables. Most major variables are statistically significant. Schools with students with higher standardized test scores have significantly higher graduation rates than those with lower scores. The coefficients for private schools and those with a higher percentage of females are also statistically significant at the 1% level. Schools with a larger portion of students receiving Pell Grants have lower graduation rates. Unfortunately, schools that have more students from lower socio-economic backgrounds are less successful in having their students' complete college. And schools with a higher percentage of Black students have lower graduation rates.

Table 3: OLS Regression: Dependent Variable: Graduation Rate

Independent Variables	Coefficients	Standard Error	T-Statistics
Intercept	-0.337	0.104	-3.234
LibArt	0.007	0.013	0.532
National	0.005	0.014	0.382
Private	0.076	0.018	4.292***
Urban	0-.004	0.010	-0.413
Suburban	0.007	0.012	0.597
IA	0.022	0.020	1.09
IN	0.009	0.021	0.45
KY	-0.074	0.020	-3.717***
MI	-0.014	0.017	-0.819
MN	0.022	0.020	1.086
MO	-0.0007	0.019	-0.351
OH	-0.004	0.017	-0.221
WI	0.027	0.019	1.472*
PerClsU20	-0.049	0.041	-1.195
AveACT	0.025	0.004	6.411***
Fresh10	0.107	0.068	1.567*
PerFemale	0.139	0.041	3.398***
OutState	-0.060	0.027	-2.192***
Interntl	0.089	0.119	0.749
Black	-0.146	0.067	-2.174***
Asian	-0.273	0.177	-1.539
Hispsc	0.093	0.096	0.967
Other	-0.028	0.100	-0.28
PerFTFac	0.093	0.048	1.942***
SalAssocProf	0.0000001	0.000	1.478**
Students	0.000003	0.000	3.366***
PerLiveOn	0.076	0.028	2.733***
PerStudFT	0.001	0.000	2.211***
PellGrant	-0.001	0.000	-2.856***
PerCapEndow	-0.00000006	0.000	-0.11
AlmnGivRt	0.158	0.081	1.958***
Number of Obs	283		
R-Squared	0.863		
Adjusted R-Squared	0.848		
F-Statistic	50.9		

*This table shows the regression: Graduation rate_i = β₀ + β₁ Time Invariant + β₂ Quality + β₃ Demographic + β₄ Faculty + β₅ Students + β₆ Financial + ε_i. The levels of statistical significance are: * at the 10% level, ** at the 5% level, *** at the 1% level. The excluded categories for the type of school is Regional (keeping Liberal Arts and National). Illinois is excluded for the different states. For ethnicity: Black, Asian and Hispanic are included and White is not.*

For this data set focusing on the Mid-West, colleges in Kentucky have lower graduation rates than those in other states. Intuitively this makes sense, while almost all institutions of higher education have seen their budgets cut over the previous decade; schools in some states such as Kentucky have had theirs cut more than others have. It is not surprising that less spending would contribute to lower graduation rates, compared to schools in other states.

I also have unique variables in this data set that are not in other studies. The signs of the coefficients and the levels of statistical significance conform to expectations. The coefficient for the percentage of students living on campus is positive and significant at the one percent level, implying that the more ‘connected’ students are to their schools the higher the graduation rate, all else constant. The alumni giving coefficient is the same, students that have a better experience likely graduate at higher rates and then tend to give more back to the school.

The F-statistic is about 51 and the R squared for the simple Ordinary Least Squares (OLS) regression is 0.863, (the Adjusted R squared is 0.850) showing that most of the variance in graduation rates is explained by these thirty-four variables. R squared in other studies was around 0.65, with the smallest 0.44 (Gansmer-Topf and Schuh, 2006) to the highest 0.85 (Scott, et al., 2006). My data is at the high end of including relevant variables impacting graduation rates.

Many other regressions were estimated. A Stepwise Regression suggested including almost the exact same variables, but I kept the regression in Table 3 because it conforms more with existing literature. Econometricians (Greene, 1993) have determined that many monetary variables are better estimated in logarithmic form so I transformed those such as per capita endowment. In addition, the percentage of freshman in the top 10 percent of their high school class was found to be increasing at a decreasing rate. Thus, the new regression included it both unchanged and also squared.

Table 4 shows a regression with some variables in logarithmic form with the other variables not transformed. Results are almost identical to those in Table 3. The F-statistic is slightly higher, 53.7 compared to 50.9, again illustrating that the model does a very good job explaining graduation rates. R squared is 0.866 (Adjusted R Squared is 0.848), compared to 0.863 and 0.850, showing that a large percentage of the variation in graduation rates in Midwestern schools is explained by the independent variables.

The results in Table 4 are comparable to those in Table 3. There were some trivial changes in a few coefficients, such as the sign on the Kentucky variable going from -0.074 to -0.093 but the level of statistical significance did not vary. Thus, the results, such as public schools having a lower graduation rates than private and demographic, such as schools with a lower percentage of males having a lower graduation rates, is the same as previous studies.

All of the current quantitative research on graduations rates at institutions of higher education stops after analyzing and interpreting the results of the regression. For example, Anstine and Seidman (2017) find that financial variables are more important than social variables, such as the percentage of students in Fraternities and Sororities in influencing graduation rates. This paper extends the analysis looking at the differences between actual and predicted graduation rates and looking at potential reasons why in the next section.

Table 4: OLS Regression: Dependent Variable: Graduation Rate, Some Variables Transformed

Independent Variables	Coefficients	Standard Error	T-Statistics
Intercept	0-.337	0.104	
LibArt	0.015	0.013	1.094
National	0.001	0.014	0.099
Private	0.083	0.019	4.317***
Urban	-0.010	0.010	-0.940
Suburban	0.008	0.012	0.646
IA	0.015	0.020	0.758
IN	-0.001	0.020	-0.033
KY	-0.093	0.020	-4.758***
MI	-0.018	0.017	-1.068
MN	0.021	0.019	1.060
MO	-0.023	0.019	-1.213
OH	-0.011	0.017	-0.643
WI	0.018	0.019	0.985
PerClsU20	-0.005	0.045	-0.105
AveACT	0.025	0.004	6.569***
Fresh10	0.265	0.113	2.345***
Fr10Sqrd	-0.207	0.106	-1.961***
PerFemale	0.080	0.042	1.916***
OutState	-0.040	0.027	-1.455**
Interntl	0.017	0.119	0.142
Black	-0.136	0.065	-2.086***
Asian	-0.237	0.176	-1.346
Hispsc	0.110	0.095	1.156
Other	-0.027	0.099	-0.271
PerFTFac	0.103	0.047	2.183***
LnStudents	0.044	0.010	4.237***
PerLiveOn	0.060	0.027	2.194***
LnEndow	0.005	0.006	0.927
LnGiving	0.040	0.009	4.267***
LnSalary	0.017	0.040	0.423
Number of Obs	283		
R-Squared	0.866		
Adjusted R-Squared	0.850		
F-Statistic	53.7		

This table shows a regression with the graduation rate as the dependent variable as a function of independent variables.

Graduation rate_i = β₀ + β₁ Time Invariant + β₂ Quality + β₃ Demographic + β₄ Faculty + β₅ Students + β₆ Financial + ε_i

*Some variables have been transformed. The percentage of students in the top 10 percent has been squared. The number of students, per capita endowment, alumni giving and faculty salaries have been transformed into logarithms. For the qualitative variables with more than two options the excluded categories are as follows: a: Regional, b: Rural, c: Illinois, d: White. The levels of statistical significance are: * at the 10% level, ** at the 5% level, *** at the 1% level.*

Qualitative Analysis

The R squared for the first (OLS) regression is 0.863, and 0.866 for the second regression. Because they are so similar, I will use the first one for the qualitative analysis. Theoretically, an R-squared of 0.863 shows that 86.3 percent of graduation rates is explained by the thirty-four explanatory variables in the model. Hence, 13.7 percent is not explained.

Forty schools are within one percentage of their predicted graduation rate being equal to their actual graduation rate. There do not seem to be any major differences between these schools. There are large, public, Research 1 (University of Illinois- Urbana, slightly above and University of Missouri, slightly below and small, private ones (Loras College in Iowa and Judson in Illinois).

In examining the characteristics of schools with largest positive difference between their predicted and actual graduation rates, two stand out, the University of Wisconsin-Madison and Perdue University are well known, 'name' universities. Otherwise, there does not seem to be any 'magic bullet' for schools with higher than predicted graduation rates. They are small, medium and large, public and private in rural, suburban and urban areas. Looking at their academics there is nothing unusual such as offering four-credit hour classes instead of three.

The main characteristic of the ten schools with actual graduation rates significantly below their predicted ones is that none of them is a 'name' school with national recognition (which is not to imply that they are not good schools). In addition, through size of school is included in the regression they are smaller than most other schools with some enrollments under 1,000 students. There is likely some other unobservable or unmeasurable characteristics related to size.

One potential factor that may lead to lower graduation rates is that some of the schools seem to an 'identity crisis.' Brescia University was founded as a Junior College for Women, but then later transformed into a coeducational, four-year school and did not become a university until 1998. Perhaps, male, female or both, students are not comfortable with the schools culture after they arrive on campus. Lindsey Wilson College is listed as a four-year, private institution, but in addition to offering Bachelor's Degrees it also offers two-year associate degrees and Ph.D.'s. Possibly by trying to be all things to different students with different needs and goals the typical four-year degree students are not given the necessary support.

Large, well known schools with huge endowments and extremely stable finances are getting more applications and graduate students at high rates, there are many institutions at the other end who are struggling. Northland College is one of these, thus the Higher Learning Commission has required a financial recovery plan for it. This likely played a large role in it being one the schools with its graduation rate being significantly below its predicted one.

While this research examines some unmeasurable characteristics that likely impact graduation rates, its biggest shortcoming is not addressing all of these potential influences. With the increase in the number of students with depression and other mental illnesses attending college, it is very likely that variables such as the resources provided for mental health impact retention and graduation. In addition, I have not looked at factors such as the existence of, and quality of First-Year seminar classes that may increase retention and graduation rates.

CONCLUSION

The goal of this paper was to extend previous literature by combining both quantitative and qualitative analysis to determine why some colleges and universities have significantly higher or lower graduation rates than predicted. Data is from The Integrated Post-Secondary Education Data System (IPEDS) and calculated directly by the author, such as the location of each school.

The first part of this research uses regression analysis on dozens of variables, controlling for different characteristics in schools to determine what variables impact graduation rates. Standardized test scores, ethnicity, the type of school, financial viability and socio-economic status predict the majority of differences on graduation rates between institutions of higher education.

The second section on methodology extends the analysis by examining specific schools with either very high or low actual graduation rates compared to their predicted graduation rates. In higher education, in many situations it seems as though the rich are getting richer (Ivy League schools) and the poor, poorer (small regional schools). The qualitative analysis section provides support for this. Some large, well know schools have higher graduation rates than predicted and smaller schools with few resources have lower graduation rates than predicted.

Educators could use this information in the following manner. They would start controlling for differences using econometrics and then focus on specific schools that underperform compared to their peer institutions. Specifically, many institutions apply High Impact Practices (HIPs), but it is likely that there are differences in the how they are implemented and resources provided for them. Qualitative analysis could look at differences, for example, in Professional Advising, after controlling for other relevant variables.

A limitation of this research is that are likely more variables that impact graduation rates that are not in the data set. The qualitative analysis tries to address this, however there are likely other influences determining graduation rates that the author did not see or was unobservable. Thus future research could examine more specific factors at different schools that influence why some schools have significantly lower graduation rates than predicted.

APPENDIX

This data looks at 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

Anstine, Jeff and Megan Seidman. (2017) "Graduation Rates at U.S. Colleges and Universities In The Midwest," *Business Education & Accreditation* vol. 9(1) p. 43-54.

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.

Einbinder, Susan. (2018) “A Process and Outcome Evaluation of a On-Semester Faculty Learning Community: How Universities Can Help Faculty Implement High Impact Practices,” *Insight: A Journal of Scholarly Teaching* vol. 13 p. 40-58.

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.

Greene, William. (1993) "Econometric Analysis," Macmillan Publishing Company, Second edition.

Hester, Brian and Terry Ishitani. (2018) "Institutional expenditures and State Economic Factors Influencing 2012-2014 Public University Graduation Rates." *Planning for Higher Education Journal* vol. 46(4) p. 41-47.

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.

Mulhere, Kaitlin and Julia Glum. (September, 2018) "The Best Colleges for Your Money." *Money Magazine*, p. 46-55.

National Center for Education Statistics, The Integrated Post-Secondary Education Data System (IPEDS), 2016, retrieved from <http://nces.ed.gov/ipeds/>

Ober, David, Beekman, John and Rebecca Pierce. (2018) "Analyzing Four-Year Public University and Two-Year College Graduation Rates," *Journal of Education and Training Studies* vol. 6(4) p 221-247.

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.

Scott, Paul. (2017) "Student Perception of College Value: Opportunities for Future Research." *Scholarship of Teaching and Learning in Psychology* vol. 5(4) p. 299-303.

Tugend, Alina. (July 19, 2019) "What Does It Mean to Be an Efficient University?" *The Chronicle of Higher Education*. Vol 65 issue 37, p. A9-A11.

ACKNOWLEDGEMENTS

I would like to thank the two anonymous reviewers and the editor for helpful comments that improved this paper. Any remaining errors is the authors'.

BIOGRAPHY

Jeff Anstine is a Professor of Quantitative Business and Environmental Management. He teaches a wide range of classes from statistics to environmental economics. His research examines topics in higher education and environmental issues. He can be reached at North Central College in Naperville IL.