FACTORS AFFECTING DROPOUT RATES IN PUBLIC UNIVERSITIES IN THE US

12 Dec 2013

Minahil Amin
Mai Le
Caryl Marten
Vimal Nair

EXECUTIVE SUMMARY

This report aims to uncover possible explanations as to what causes college dropout rates to differ at one hundred and four public universities across the United States. Given that college completion rates serve as a measure of future workforce productivity and only 39.6% of students who matriculate to bachelor's degree programs graduate with a degree within five years (ACT, 2010), policymakers would be wise to devote more attention to the issue of attrition in higher education. By incorporating the findings of Turner, Lovenheim, and Bound (2009) and the Katherine Aloisi, Colleen McEnearney, and Giana Solomon's IPA entitled "Improving Retention and Completion Rates at Maryland Four-Year Post-Secondary Institutions" into our own hypotheses, we elected to examine the effect of five institutional and financial variables - percentage of minority students, number of employees, academic support spending, grant aid, and tuition - on college dropout rates. To perform this analysis, we conducted t-tests, correlation, OLS, and regression tests while also checking for heteroscedasticity and multicollinearity. The analysis ultimately found that larger numbers of university employees, higher rates of academic support spending, and greater amounts of grant aid all had statistically significant relationships with lower college dropout rates.

INTRODUCTION

Inspired by Katherine Aloisi, Colleen McEnearney, and Giana Solomon's IPA entitled "Improving Retention and Completion Rates at Maryland Four-Year Post-Secondary Institutions" (hereby referred to as Aloisi et al.), the motivation of this report is to identify the causes of high college dropout rates at one hundred four public universities across the nation. As the world moves toward a service economy, brainpower becomes the most valuable commodity. If the United States wishes to maintain its position as a global economic power, it must continue to support its citizens' pursuits of higher education. In order to help achieve this, this report identifies the key factors which most impede these educational endeavors and should therefore be the focus of policies aimed at lowering college dropout rates.

In order to conduct this analysis, we used Aloisi et al.'s finding that the affordability of higher education plays a critical role in disenrolling from college as a foundation for our own hypothesis. In addition to conducting an analysis of financial factors - average grant aid and tuition - in order to affirm or disprove the IPA's conclusion, we expanded our pool of independent variables to include institutional factors. After an initial regression to determine significance, we were left with three institutional factors to assess: the percentage of enrolled students that are minorities, the number of the university's employees, and the average amount of academic support spending per full-time student. Once these five variables were identified, t-tests, correlation tests, ordinary least squares tests, and regression tests (both standalone and with controlling variables to detect direct and indirect effect) were conducted to examine if any significant relationships between the variables and dropout rates existed. Residual plots were also generated in order to identify outliers and nonlinear forms, and the hettest was employed to test for heteroscedasticity.

This report identifies the number of a school's employees, academic support spending, and grant aid as factors which have a statistically significant impact on college dropout rates. Schools with more employees, higher rates of academic support spending, and larger amounts of grant aid are likely to have lower dropout rates. While this finding affirms the Aloisi et al. conclusion that affordability is a critical factor in the abandonment of bachelor's degree programs, it also pinpoints additional institutional factors which strongly impact dropout rates. This is ultimately significant because it provides a fuller picture of why students remove themselves from school and expands the number of areas within higher education policy which can be reformed to combat stagnant college graduation rates.

PRIOR RESEARCH

In their IPA, Aloisi et al were concerned with finding ways to improve completion rates for bachelor's degree programs to increase workforce productivity and attain long term economic benefits, specifically in the state of Maryland.

Their motivation in choosing this topic was the fact that, nationwide and in Maryland, completion rates are significantly lower than enrollment rates. They found that at four-year public institutions across the US, only 39.6 percent of students who enroll in a bachelor's degree program will end up graduating in five years or less (ACT, 2010). "Student attrition hurts the individual student and the institution from which he drops out, as well as the state's endowment of human capital and economic competitiveness. Every student who fails to complete also represents lost opportunity in terms of wages, tax revenue, investments, and job creation."

After examining the literature on college education, Aloisi et al concluded that student completion is affected by such factors as academic performance, quality of social interactions and social integration, and affordability of higher education. An effective retention strategy involves targeting both academic and non-academic factors that cause students to drop out of college.

They chose to focus on financial constraints as the primary variable affecting completion rates. They found data from the College Board showing that financial factors affect completion because academically well-prepared students from low-income families have lower completion rates than academically well-prepared students from high-income families. They saw this issue as an opportunity for policymakers: "There is great potential for the financial aid system to boost enrollment and completion."

Their reasoning behind this claim was that financial aid improves retention because it provides aid to students who otherwise would not be able to finance their education. It also decreases the likelihood that students will need a job to meet their expenses, thus allowing for greater engagement in the academic and social environments of college.

Although Aloisi et al did not conduct their own analysis on financial aid and completion rates, they found additional studies showing positive correlations between amount of grant aid given to students and years of educational attainment at college (Dynarski, 1999, p. 3).

Based on this research, their ultimate policy recommendation was for the state of Maryland to design programs that reduce the time to degree and to streamline the financial aid process.

For our analysis, we chose to focus on a claim highlighted in the IPA: "Institutional factors have also been identified as playing a significant role in declining completion rates." A study by Turner, Lovenheim, and Bound (2009) found that though student preparedness and collegiate characteristics play a role in declining completion rates, institutional characteristics are more important, accounting for one quarter of the decline in completion rates.

While Aloisi et al. acknowledged this finding, they did not delve into it further and instead chose to base their policy recommendation solely off of financial factors. They backed their policy fairly well, but we think that the policy recommendation would change or expand if a thorough analysis of other factors affecting college dropout rates were done.

If our research reveals that institutional factors play a strong part in student persistence at college, then the emphasis of the policy will shift from designing programs that reduce the time to degree (in order to save students money) to adapting institutions to increase the likelihood of students receiving a degree. If, for example, we find that institutions that spend a higher amount on student services have lower dropout rates, we will encourage schools to implement more and better quality student services.

At the same time, we will reexamine the authors' main claim about the indubitable positive effect of financial aid on completion rates. If we find that colleges that have students with better financial support (through aid) have lower dropout rates, we will validate the authors' claims. If financial aid is an integral factor in completion, then the original policy recommendation to streamline the aid process is very reasonable.

If, however, we find that financial aid plays a negligible role in dropout rates, we will refute the authors' hypothesis and policy, and shift the emphasis of the policy recommendation completely to institutional factors.

RESEARCH DESIGN

As the main aim of our re-examination of this IPA was to more fully understand the role of institutional factors on dropout rates in universities, our hypotheses are as follows:

H_0: There is a no correlation between institutional factors and college dropout rates at major public universities across the US.

H_A: There is a significant correlation between institutional factors and college dropout rates at major public universities across the US.

We expanded the scope of the IPA and collected data on major public universities across the country for the year 2011. The variables that we consider integral to examining the relationship between institutional factors and drop-out rate are: academic spending per fully enrolled student (AcadSupport2011), student services spending per fully enrolled student (StudServ2011), student-to- faculty ratio (StudFacRatio2011), number of employees (Employees2011), and percentage of undergraduates that are white (EnrollWhite2011). We also wanted to reexamine the original assertion of the IPA - that financial factors were the main influencers of dropout rates, so we included data for tuition (Tuition1112), average grant aid per student (AvgGrantAid1112), and percentage of students at the college receiving grant aid (PercentGrantAid1112).

We created a variable for dropout rate (DropOutRate2011) by subtracting the graduation rate from 100. We ran an initial regression with the dropout rate variable and our predictor variables to ascertain which of our predictor variables were significant. We found that AcadSupport2011, Employees2011, EnrollWhite2011, Tuition1112, and AvgGrantAid1112 were statistically significant.

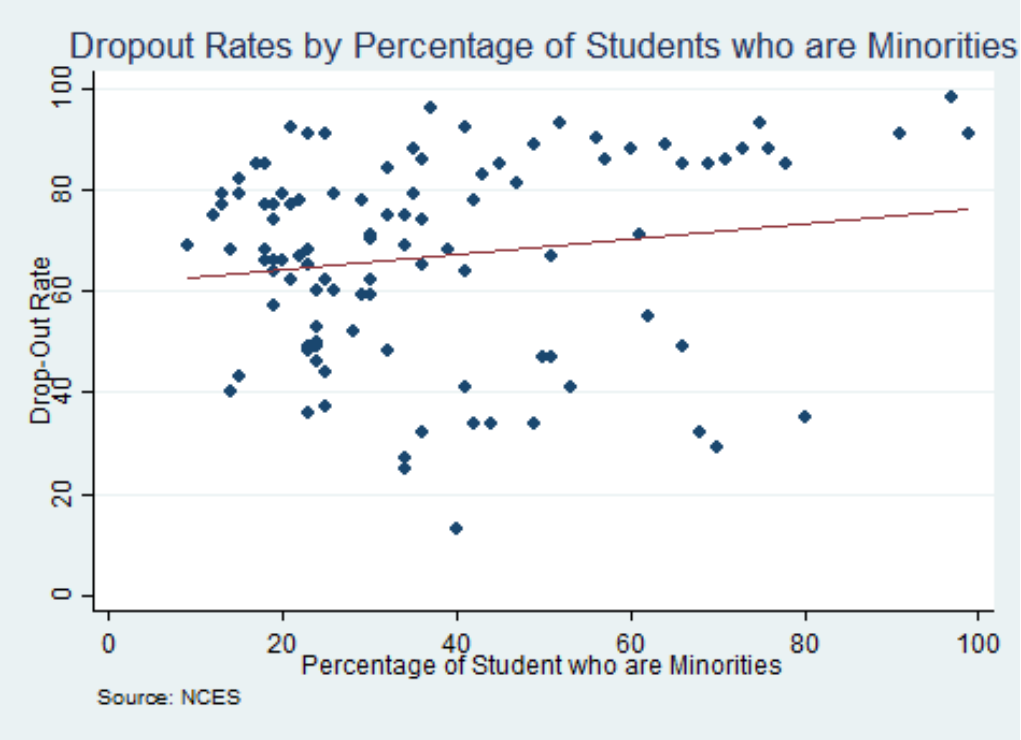
On each of these variables, we ran regressions against the dropout rate variable, making sure to also check for robustness by carrying out tests for outliers, nonlinear forms, heteroscedasticity, multicollinearity, and effects of omitted variables. To deal with outliers, we created twoway scatter plots of dropout rate versus the dependent variable in question. We also plotted plots of residuals versus the dependent variable. If we found outliers, we either deleted the outliers or transformed the variable with outliers to bring it closer to the median. We also used the visualization of the data to check for nonlinear forms and adjust the regression to include higher order effects. For heteroskedasticity, we ran the hettest after each regression and made sure that each regression command specifies the “robust” option. For multicollinearity, we checked the variance inflation factor via the the vif command. Our rule of thumb for the VIF was to investigate variables whose VIF values were greater than 10. Fortunately, all of our predictor variables had VIF values significantly below 10.

To help ensure that the relationships we find are not spurious, we will add in control variables in the form of variables related to students’ financials that might help explain dropout rates. When examining the relationships between institutional variables and dropout rates, we used tuition as our control. In examining the relationship between tuition and dropout rates, we used average grant aid as our control.

ANALYSIS

Institutional Variable #1: Percentage of Enrolled Students that are Minorities

To examine how minority enrollment affected dropout rates, we created the variable MinorityEnroll2011 by taking the variable for percentage of students that are white and subtracting it from 100 (100 - EnrollWhite2011). A twoway scatter plot reveals a slightly positive trend between percentage of enrolled students who are minorities and drop-out rate.



A test for correlation confirms the positive trend between dropout rates and Minority enrollment with a correlation coefficient of 0.1560, but with a p-value of 0.1137 this trend is not significant at a 95% confidence interval. If we use a t-test with a dichotomous version of the percentage of minority enrollment (where a value of 1 indicates that the percentage of students who classified themselves as minorities was greater than 30%), we discover that schools with at least 30% minority students have a 1.92% higher dropout rate. The null hypothesis for this t-test would be that there is no difference in means between the two dichotomous variables (with the alternative hypothesis saying that there is a difference) but this null hypothesis seems to be rebuffed by the fact that schools with greater than 30% minority students have a higher dropout rate than schools with less than 30% minority students. However when we look at the p-value for this difference, it is only 0.6902 which means that this difference is not significant at the 95% confidence interval and thus we cannot reject the null hypothesis in this instance.

The residual plot for this minority enrollment variable show a large amount of variance between schools with the average residual score being approximately 20 in either direction. In addition to that the residual plot shows a couple of outliers, especially when the percentage of minority students exceeds 30%. These residual plots indicate that the model is unable to be very accurate due to the high variance.

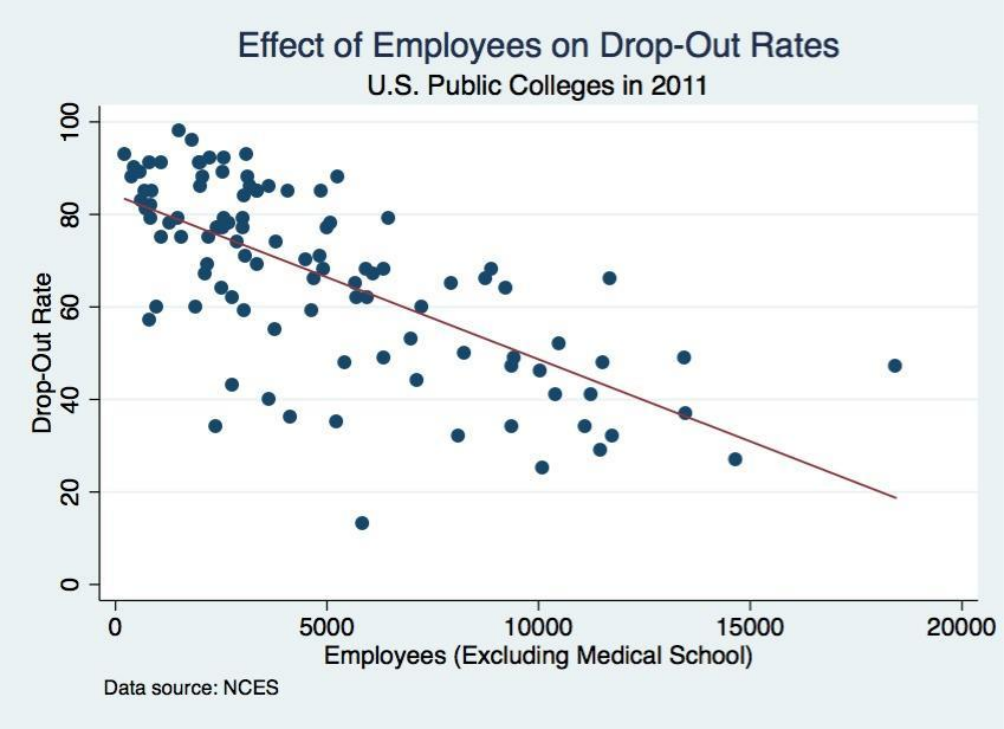
For each percentage increase of minority enrollment, the dropout rate increases by 0.1496%. Meaning that if the percentage of minority students at a school increased by 10%, then this regression predicts that the dropout rate would increase by 1.496%. Once again, the regression coefficient is not significant on a 95% confidence level since the p-value is 0.122.

If we control for tuition and average grant aid, then the effect of minority enrollment actually increases to 0.2548 and the p-value drops to below 0.0001 which means that this correlation coefficient is significant with at least 95% confidence.

After conducting all of the tests on the minority enrollment variable, we cannot say with 95% confidence that the percentage of students who are minorities has an impact on dropout rates at the Universities we looked at. These findings confirm the null hypothesis that this institutional variable does not have an impact on dropout rates. The only possible rebuttal to this argument is that when we controlled for average grant aid and tuition, the regression coefficient for Minority Enrollment because statistically significant and thus could rebuff the null hypothesis. However, it would be hard to accept this argument since the t-test, residuals, and pwcorr all indicate otherwise.

Institutional Variable #2: Number of Institution Employees

We also examined how the number of employees at an institution might affect dropout rates. A two-way scatter plot reveals a negative trend. Each additional employee creates a corresponding decrease in dropout rates.



A correlation test yields a statistically significant correlation coefficient of -0.6814 for this relationship. After transforming the Employees variable into a dichotomous form (where a value of 1 indicates that the number of employees in the institution is above 3368, and where a value of 0 indicates that the number of employees is at or below 3368), a t-test results in a statistically significant difference of means of 23.76. The p-value is smaller than 0.0001. Therefore, we can accept the alternative hypothesis (that there is a difference in the two means between the dichotomous variables created--as opposed to the null in which there is no difference) at the 99% confidence level. Colleges with below-median (3368) employee numbers had higher dropout rates than colleges above the median.

A robust regression, with checks for OLS assumptions, confirms the statistically significant relationship between employees and dropout rates. For each each additional employee, dropout rates fall by .0035 percentage points. The p-value is less than 0.0001, therefore the regression coefficient is significant at a 99% confidence level.

The residual plot for the employment variable shows a positively-sloped trend. This may indicate that a linear regression line is not the best fit for regression equation because of heteroscedasticity. After conducting a test for heteroscedasticity ("hettest" in Stata), the chi-value that emerged was 0.46 (with a large p-value of .499). This chi-value is relatively small, indicating that heteroscedasticity (the variability of the variable is unequal across the range of values) is unlikely to be problematic in our analysis. Therefore, the regression model we are using is acceptable.

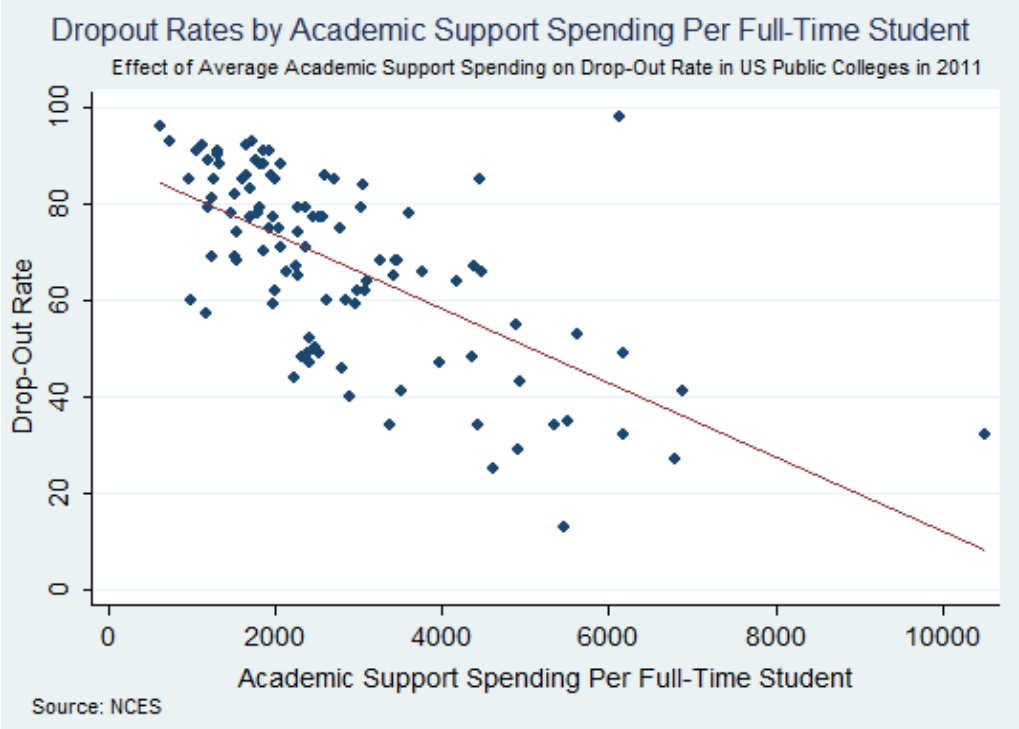
When we control for tuition, the regression equation indicates that with every increase of adding one additional employee to the institution, dropout rates will fall by 0.0022 percent.

Overall, several tests come to the same conclusion that the fewer employees an institution employs, the more likely it is to have higher dropout rates. Although the aforementioned tests are not enough of an analysis to prove causation, they do indicate an association between dropout rates and the number of employees at an institution.

Institutional Variable #3: Average Academic Support Spending per Full-Time Student

Continuing our effort to examine the effect of institutional factors, we also explored how academic support spending impacts dropout rates. AcadSupport2011, the variable which measures average academic support spending per full-time student, measures spending on programs and resources which contribute to students' educational success but are considered separate from classroom experience. This includes spending on libraries, academic deans and administration, information technology resources available to students, and student and faculty personal development programs which aim to improve academic effectiveness.

Our research hypothesis for this variable was that schools with higher rates of academic support spending would have lower dropout rates, since students will have more resources to turn to for support before dropping out. This hypothesis was supported by the two-way scatter plot which was generated between these two variables that can be seen below. The scatter plot depicts a negative trend: as average academic support spending increases, dropout rates decrease.



The correlation test confirmed this negative relationship, yielding a statistically significant (p-value = 0.0000) coefficient of -0.6492.

In order to conduct a t-test to further test the relationship, 2011 academic support spending was transformed into a dichotomous variable, with a value of 1 indicating average academic support spending per full-time student of more than the median $2,364, and a value of 0 indicating average academic support spending of $2,364 or less. The t-test then showed a difference in the means of the two groups which further affirmed our hypothesis; the average dropout rate of schools whose academic support spending is above the median is 22.93% less than the average dropout rate of schools with below-median academic support spending. Since the p-value for this test was 0.0000, we can reject the null hypothesis that suggested there would be no difference between the two means at the 95% confidence level and accept that the difference between below-median and above-median schools is statistically significant.

Furthermore, a robust regression with checks for OLS assumptions producing a p-value of 0.0000 again confirmed the statistically significant negative relationship between academic support spending and dropout rates. Each dollar increase in academic support spending is accompanied by a .0077% decrease in the dropout rate. Given that an $1000 increase in academic support spending would result in a corresponding 7.7% decrease in a school's dropout rate, this impact of this academic support spending on dropout rates is clearly substantively significant. The residual plots for this regression did not show any clearly defined trends, although there was one outlier with a high residual score of 56.09. Given that this is only one data point in a sample size of 100, it should not have impacted our model.

Rerunning the regression test while controlling for tuition reduced the effect of academic support spending on dropout rates by nearly half, to a dropout rate decrease of 0.0043 for every $1 increase in academic support spending. In order to delve deeper into this analysis, we calculated the indirect and direct effects of average academic support spending on dropout rates. Given that the total effect of academic support spending on dropout rates is -0.0077108 and the fact that total effect = direct effect + indirect effect, we found that

$$\text{Total effect} = -0.0077108 = -0.0042885 + (-0.0036165)(.9463095)$$

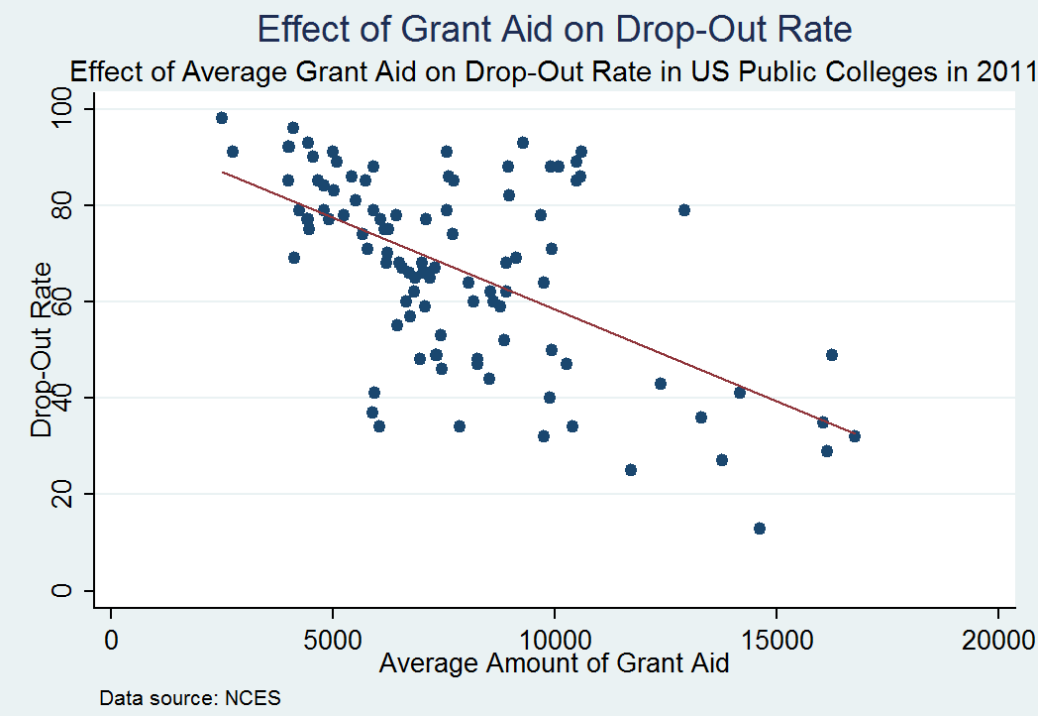
This demonstrates that the direct effect of academic support spending on dropout rates constitutes 55.62% of the total effect. Given that this is still a majority of academic support spending's total effect on dropout rates, we can still conclude that the effect of academic support spending is still substantively significant.

Since all of the tests conducted to examine the impact of academic support spending on dropout rates demonstrated relationships which are statistically significant with p-values of less than

0.0001, we can say with 95% confidence that average academic support spending has an impact on dropout rates at the 100 universities studied in this report.

Financial Variables: Average Grant Aid and Tuition

The final variable we are examining is the average amount of grant aid per student at a college in 2011 (AvgGrantAid1112). The purpose of examining this variable is to reexamine the authors' original claim – that financial aid significantly lowers dropout rates. Running a two-way scatter reveals that there is a relatively strong negative correlation between grant aid and dropout rates – that is, each additional dollar of average grant aid has a corresponding decrease in dropout rate.



The results of our other tests further substantiate this finding. A test of correlation yields a statistically significant correlation coefficient of -0.5834 for this relationship. After transforming the average grant aid variable into a dichotomous form, we ran a t-test, The null hypothesis for this t-test would be that there is no difference in means between the two dichotomous variables (with the alternative hypothesis saying that there is a difference) but this null hypothesis seems to be rebuffed by the fact that there is a statistically significant (at a 95% confidence level) difference of means of 15.0 percentage points in dropout rates; colleges whose students receive below-median grant aid had significantly higher dropout rates than colleges above the median.

A robust regression, with checks for OLS assumptions, confirms the statistically significant relationship between average grant aid and dropout rates. For each dollar increase in grant aid, dropout rates fall by .0038 percentage points. Thus, for every $1000 increase in grant aid,

completion increases by 3.8%. The residual plots for this regression don't show any trends, outliers, or fanning. They suggest that the model is working as well as it can.

Since all of the tests conducted to examine the impact of academic support spending on dropout rates demonstrated relationships which are statistically significant with p-values of less than 0.0001, we can say with 95% confidence that average amount of grant aid per student has an impact on dropout rates at the 100 universities studied in this report.

When we control for tuition, the strength of this relationship is halved. While still statistically significant, every $1000 increase in grant aid yields only a 1.7% fall in dropout rates.

To further examine this result, we calculated the indirect and direct effects of average grant aid on dropout rates. Total effect = direct effect and indirect effect

$= \widehat{\beta_1} + \widehat{\beta_2}\widehat{\beta_{X_2X_1}} = -.001689 + (-.00405)(0.5237).$ It turns out that the direct effect of grant aid on dropout rates accounts for 44.3% of the total effect. The indirect effect implies that the amount of average grant aid students at a university get affects the tuition the university charges, which in turn helps decrease dropout rates.

This finding, although not relevant to our null hypothesis, *is* integral to our analysis as it helps support the hypothesis of Aloisi et al. A $1000 increase in average aid to students at a university is directly correlated with a 1.7 percent decline in dropout rates. If you take into account the theory that grant aid acts as a tuition "cushion" and allows universities to charge higher tuition without narrowing its pool of lower-income applicants, the argument for increased amounts of grant aid is even stronger. These findings are substantively significant. The average enrollment at public universities across the US in 2011 was 18,743.36. A 1.7 or 3.8 percent decrease in dropout means that 319 or 712 more students at each university will graduate, get better jobs, and be more productive contributors to the economy.

Thus, Aloisi et al's policy recommendation of streamlining the financial aid process so more students can get aid is astute.

CAVEATS AND CONCERNS

The major caveat to this data is the fact that although all of our data came from one source (NCES), the data itself is only available from 2011. This means that this is not the most recent data available and the findings from these data may not be relevant in understanding what leads to an increased dropout rate today. This data can only provide a snapshot of how institutional and financial variables affected dropout rates in 2011 for these 104 schools.

When deciding on how to define dropout rates, we used the percentage of students who graduated with a bachelor's degree within four years and then subtracted that from 100%. This means that within the students we are classifying as dropouts are students who may have taken 5 or 6 years to complete their degree. In the original study, they calculated dropout rates by looking at those that were able to finish their degree within 5 years so our findings may not be necessarily relevant to their analysis since we have a much larger pool we are considering "dropouts." This issue limits our ability to relate our findings to those of Turner, Lovenheim, and Bound (2009).

Finally, for the white enrollment/ minority enrollment variables, we can not be certain that they are accurate representations of the racial demographics of the Universities because students have no obligation to indicate their race to the University. This means that when we used the white enrollment variable, we may have not been receiving the accurate proportion of students that would necessarily be classified as "white."The possible discrepancy in proportion could have skewed the findings on how racial demographics affected dropout rates.

IMPLICATIONS AND CONCLUSIONS

Through the use of t-tests, correlation tests, regression tests, the analysis conducted in this report identified three variables - one institutional, and one financial as followed from the findings in Aloisi et al. - as having statistically significant relationships with college dropout rates. More school employees, academic support spending, and grant aid were all correlated with lower dropout rates.

Minority enrollment is inconclusive. Except for when we run the regression while controlling for tuition and average great aid, all of the other tests were not found to be significant on the 95% confidence interval.

In terms of relative substantive significance, when we control for tuition: every $1000 increase in academic support spending would result in a corresponding 4.3% decrease in a school's dropout rate, every $1000 increase in average grant aid per student would result in a corresponding 1.7% decrease in a school's dropout rate, and every additional employee would result in a corresponding 0.0022% decrease in a school's dropout rate. One key takeaway from these findings is that academic support spending has a relatively large impact on dropout rates, indicating that schools that invest in ensuring that their students can cope with the academic requirements of a degree have more students graduating on time with that degree. This finding also affects the original policy recommendation in the IPA.

Had one attempted to guide higher education policy toward lowering dropout rates after reading the original report which inspired our own investigation, the reforms would have been geared completely toward students' individual costs of pursuing higher education, as Aloisi et al. made recommendations based solely on the issue of enhancing the affordability of college. While our examination helps to confirm that their recommendations of reducing degree times and streamlining the financial aid process are reasonable, the fact of the matter is that the fragile state of the economy means that financial resources are limited for both families and the sources of educational grants. Our discovery concerning institutional factors' impact on dropout rates provides new, more fiscally feasible outlets that public universities can employ in order to address dropout rates at a time when budgets are tight. Under the Aloisi et al. analysis, the only methods to increase retention rates involve a serious expenditure by the universities, whether they be in the form of grants or expedited graduation schedules (which result in lost tuition revenue). The institutional factor findings mean that universities can fund new academic support resources or hire more employees in order to lower dropout rates by either reallocating their budget or marginally increasing expenditures.

Appendix

	Description	Mean	Std. Dev	Min	Median	Max	Source
GradRate2011	% under-graduate students; graduate in 4 years.	33.115	19.624	2	31	87	NCES nces.ed.gov
DropOutRate2011	Calculated by finding 100-GradRate2011	66.885	19.624	13	69	98	Calculated
AcadSupport2011	Academic support expenses per FTE	2791.835	1632.425	612	2364	10504	NCES
Employees2011	Number of institution employees (excluding medical school).	4823.886	3777.04	219	3368	18442	NCES
EnrollWhite2011	% under-graduate enrollment that are White.	63.124	20.4402	1	70	91	NCES
Tuition1112	Total tuition per student	8253.394	2754.189	4006	7440	15984	NCES
AvgGrantAid1112	Average amount of federal, state, local or institutional grant aid received	7734.543	3009.576	2504	7100	16760	NCES
PercentGrantAid-1112	% of under-graduate	59.467	11.410	40	58	89	NCES

	students receiving any grant aid, 2011-2012						
Enrollment2011	Number of under-graduate students enrolled	18743.36	9894.74	2076	19345	58404	NCES
Tuitiondichot	Dichotomous variable for tuition. Above median=1, below=0	.457	.501	0	0	1	Calculated
AvgGrantAidDichot	Dichotomous variable for Average Grant Aid. Above median=1, below=0	.495	.502	0	0	1	Calculated
MinEnrollDichot	Dichotomous variable for Minority Enrollment. Above median=1, below=0	.495	.502	0	0	1	Calculated
EmployeesDichot	Dichotomous variable for employees. Above median=1, below=0	.495	.502	0	0	1	Calculated
AcadSupportDichot	Dichotomous variable for Academic Support. Above median=1, below=0	.505	.502	0	1	1	Calculated

DO-FILE

```
/*****************************************************************************
**
                        PPOL 6150 Final Exam
                  438189124; 103392479; 907234780; 568575908
                              12 December 2012
                  Factors Affecting Drop-out Rates in Public Universities
                              Across the United States
*****************************************************************************
*/

use "C:\Users\Mai\Documents\UVa\Fall 2013\Research Methods & Data Analysis
I\Final\UniversityData.dta"

/*getting rid of excess variables*/
drop AcadSupport2010
drop AvgGrantAid0910
drop AvgGrantAid1011
drop Employees2010
drop EnrollWhite2010
drop Enrollment2010
drop Est_undergradenroll2012
drop FinAid0910
drop FinAid1011
drop HighestDegree2012
drop OpRev0910
drop OpRev1011
drop StudFacRatio10
drop StudServ2010
drop Tuition0910
drop Tuition1011
drop Tuition1213

/*creating a variable for drop out rates*/
gen DropOutRate2011=100-GradRate2011

#delimit ;
lab var DropOutRate2011
```

```
"Percentage of Students who do not graduate within four years"
;
#delimit cr

/*running initial regression for factors affecting drop out rates in 2011*/
#delimit ;
reg DropOutRate2011 AcadSupport2011 StudServ2011 Employees2011 EnrollWhite2011
 StudFacRatio11 Tuition1112 AvgGrantAid1112 PercentGrantAid1112, robust
 ;
#delimit cr

vif

#delimit ;
reg DropOutRate2011 AcadSupport2011 StudServ2011 Employees2011 EnrollWhite2011
 StudFacRatio11, robust
 ;
#delimit cr

reg DropOutRate2011 Tuition1112 AvgGrantAid1112 PercentGrantAid1112, robust

/**********DECIDING ON THE CONTROLS**********/

/*does aid affect tuition?
does tuition affect institutional factors?
does aid directly affect institutional factors?*/

reg AcadSupport2011 AvgGrantAid1112, robust
reg AcadSupport2011 AvgGrantAid1112 Tuition1112, robust
reg Tuition1112 AvgGrantAid1112, robust

/*Does the number of employees affect aid?
Does the number of employees effect tuition or viceversa?*/

reg AvgGrantAid1112 Employees2011, robust
reg Tuition1112 Employees2011, robust


/**********
*
                INSTITUTIONAL VARIABLES
```

```
**********************************************************************************
/

/******************** VARIABLE 1- EnrollWhite2011 ****************************/

*Creating Minority Enrollment Variable

gen MinorityEnroll2011=(100-EnrollWhite2011)
lab var MinorityEnroll2011 "Percentage of Students who are Minorities, 2011"

/* Decide on Controls: Does aid affect minority enrollment?
does tuition affect minority enrollment?*/

reg AvgGrantAid1112 MinorityEnroll2011, robust
reg Tuition1112 MinorityEnroll2011, robust

/*Scatter plot of EnrollWhite2011

#delimit ;
twoway (scatter DropOutRate2011 MinorityEnroll2011)
(lfit DropOutRate2011 MinorityEnroll2011),
ylabel(0 20 40 60 80 100, grid) legend(off)
title(Dropout Rates by Percentage of Students who are Minorities)
note("Source: NCES")
ytitle("Drop-Out Rate")
xtitle("Percentage of Student who are Minorities")
; #delimit cr
*/


*Correlation
pwcorr DropOutRate2011 MinorityEnroll2011, star(0.05) sig

*T-test
tabstat MinorityEnroll2011, stat(p50)
gen MinEnrollDichot=0
lab var MinEnrollDichot "Dichotomous Variable for Minority Enrollment"
replace MinEnrollDichot=1 if MinorityEnroll2011>30

ttest DropOutRate2011, by (MinEnrollDichot)
```

```
*OLS

*basic regression
reg DropOutRate2011 MinorityEnroll2011, robust

*examining residuals
predict predDORMinEnroll
predict resDORMinEnroll, res

*residuals versus predicted tuition values plot
rvfplot
*residuals versus actual tuition values plot
rvpplot MinorityEnroll2011
*residuals versus drop out rate (dependent variable) plot
plot resDORMinEnroll MinorityEnroll2011

*Checking heteroskedasticity
quietly: reg DropOutRate2011 MinorityEnroll2011
hettest

*Checking for indirect effects with tuition and average grant aid as a control
reg DropOutRate2011 MinorityEnroll2011 Tuition1112 AvgGrantAid1112, robust


/********************* VARIABLE 2- Employees2011
*******************************/

/*scatter plot of Employees2011
#delimit ;
twoway
(scatter DropOutRate2011 Employees2011)
(lfit DropOutRate2011 Employees2011)
, ylabel(0 20 40 60 80 100, grid) legend(off)
  title("Effect of Employees on Drop-Out Rates")
  subtitle("U.S. Public Colleges in 2011")
  note("Data source: NCES")
  ytitle("Drop-Out Rate")
  xtitle("Employees (Excluding Medical School)")
;
#delimit cr
*/
```

```
/*correlation*/
pwcorr DropOutRate2011 Employees2011, star(0.05) sig

/*ttest*/
tabstat Employees2011, stat(p50)

gen EmployeesDichot=0
replace EmployeesDichot=1 if Employees2011>3368

ttest DropOutRate2011, by (EmployeesDichot)

/*OLS*/

/*basic regression*/
reg DropOutRate2011 Employees2011, robust

*examining residuals
predict predDOREmp2011
predict resDOREmp2011, res

*residuals versus predicted employee number plot
rvfplot
*residuals versus actual employee number plot
rvpplot Employees2011
*residuals versus drop out rate (dependent variable) plot
plot resDOREmp2011 DropOutRate2011

*checking heteroskedasticity
quietly: reg DropOutRate2011 Employees2011
hettest

*checking for indirect effects with tuition as a control
reg DropOutRate2011 Employees2011 Tuition1112, robust
reg Employees2011 Tuition1112, robust

/********** VARIABLE 3 - Academic Support **********/

/*Scatter plot of AcadSupport2011

#delimit ;
twoway (scatter DropOutRate2011 AcadSupport2011)
```

```
(lfit DropOutRate2011 AcadSupport2011),
ylabel(0 20 40 60 80 100, grid) legend(off)
title(Dropout Rates by Academic Support Spending Per Full-Time Student)
subtitle(Effect of Average Academic Support Spending on Drop-Out Rate in US Public Colleges in 2011)
note("Source: NCES")
ytitle("Drop-Out Rate")
xtitle("Academic Support Spending Per Full-Time Student")
; #delimit cr
*/

*Deciding on Control

reg AvgGrantAid1112 AcadSupport2011, robust
reg Tuition1112 AcadSupport2011, robust

*Correlation
pwcorr DropOutRate2011 AcadSupport2011, star(0.05) sig

*T-test
tabstat AcadSupport2011, stat(p50)
gen AcadSupportDichot=0
lab var AcadSupportDichot "Dichotomous Variable for Academic Support Spending"
replace AcadSupportDichot=1 if AcadSupport2011>2364

ttest DropOutRate2011, by (AcadSupportDichot)

*OLS

*basic regression
reg DropOutRate2011 AcadSupport2011, robust

*examining residuals
predict predDORAcadSupport
predict resDORAcadSupport, res

*residuals versus predicted tuition values plot
rvfplot
*residuals versus actual tuition values plot
rvpplot AcadSupport2011
*residuals versus drop out rate (dependent variable) plot
```

```
plot resDORAcadSupport AcadSupport2011

*Checking heteroskedasticity
quietly: reg DropOutRate2011 AcadSupport2011
hettest

*Checking for indirect effects with tuition as a control
reg DropOutRate2011 AcadSupport2011 Tuition1112, robust
reg Tuition1112 AcadSupport2011, robust


/*****************************************************************************
*
                    FINANCIAL VARIABLES
*****************************************************************************
/

/****************** VARIABLE 4 - Average Grant Aid ***************************/

/*scatterplot

#delimit ;
twoway
(scatter DropOutRate2011 AvgGrantAid1112)
(lfit DropOutRate2011 AvgGrantAid1112)
, ylabel(0 20 40 60 80 100, grid) legend(off)
  title("Effect of Grant Aid on Drop-Out Rate")
  subtitle("Effect of Average Grant Aid on Drop-Out Rate in US Public Colleges in 2011")
  note("Data source: NCES")
  ytitle("Drop-Out Rate")
  xtitle("Average Amount of Grant Aid")
;
#delimit cr
*/

/*correlation*/
pwcorr DropOutRate2011 AvgGrantAid1112, star(0.05) sig

*ttest
tabstat AvgGrantAid1112, stat(p50)
gen AvgGrantAidDichot=0
```

```
replace AvgGrantAidDichot=1 if AvgGrantAid1112>7100

ttest DropOutRate2011, by (AvgGrantAidDichot)

*OLS

*basic regression
reg DropOutRate2011 AvgGrantAid1112, robust

*examining residuals
predict predDORAvgAid
predict resDORAvgAid, res
*residuals versus predicted tuition values plot
rvfplot
*residuals versus actual tuition values plot
rvpplot AvgGrantAid1112
*residuals versus drop out rate (dependent variable) plot
plot resDORAvgAid DropOutRate2011

*checking heteroskedasticity
quietly: reg DropOutRate2011 AvgGrantAid1112
hettest

*checking for indirect effects with tuition as a control
reg DropOutRate2011 AvgGrantAid1112 Tuition1112, robust
reg AvgGrantAid1112 Tuition1112, robust


/****************************** VARIABLE 5 - TUITION
**************************/

/*scatterplot

#delimit ;
twoway
(scatter DropOutRate2011 Tuition1112)
(lfit DropOutRate2011 Tuition1112)
, ylabel(0 20 40 60 80 100, grid) legend(off)
  title("Effect of Tuition on Drop-Out Rate")
  subtitle("Effect of Tuition on Drop-Out Rate in US Public Colleges in 2011")
  note("Data source: NCES")
```

```
  ytitle("Drop-Out Rate")
  xtitle("Tuition")
;
#delimit cr */

*correlation
pwcorr DropOutRate2011 Tuition1112, star(0.05) sig

*ttest
tabstat Tuition1112, stat(p50)
gen Tuitiondichot=0
replace Tuitiondichot=1 if Tuition1112>7740

ttest DropOutRate2011, by (Tuitiondichot)

*OLS

*basic regression
reg DropOutRate2011 Tuition1112, robust

*examining residuals
predict predDORTuition
predict resDORTuition, res
*residuals versus predicted tuition values plot
rvfplot
*residuals versus actual tuition values plot
rvpplot Tuition1112
*residuals versus drop out rate (dependent variable) plot
plot resDORTuition DropOutRate2011

*checking heteroskedasticity
quietly: reg DropOutRate2011 Tuition1112
hettest

*checking for indirect effects with average grant aid as a control
reg DropOutRate2011 Tuition1112 AvgGrantAid1112, robust
reg Tuition1112 AvgGrantAid1112, robust


/****************************** END DO-FILE
**********************************/
```