University selectivity and student retention: Evidence from the IPEDS data

Yue Xu[1]*, Baoqing Cheng[1]

Shanghai Normal University Tainhua College, PRC

ABSTRACT

Using the IPEDS (Integrated Postsecondary Education Data System) Data, this project examines: (1) the relationship between university selectivity and student retention; (2) whether college freshmen's ACT scores mediate the relationship between University selectivity and student retention; and (3) whether average number of full-time faculty moderates the relationship between university selectivity and student retention. The results suggest that: (1) university selectivity is positively correlated with student retention; (2) college freshmen's ACT scores fully mediate the relationship between University selectivity and student retention; (3) average number of full-time faculty does not moderate the relationship between university selectivity and student retention.

Introduction

College students' attrition is a great concern in the USA as there are large numbers of students leave during their first year of study after enrolling in postsecondary education and more students leave before finishing their college studies [1]. To explore this phenomenon, a number of studies have been conducted and various factors which may have effects on student retention have been explored. For example, Chen tests the relationship between student retention and institutional attributes and finds that institutional expenditure on student services is positively related with student retention [2]; Lee, Olson, Locke, Michelson, and Odes examine the effects of college counseling services on student retention [3]; Lin, Yu, and Chen suggest that high school GPA and percentage of class rank, graduating from a larger high school, and programs including orientation and remedial English courses, on-campus jobs, and on-campus residence are positive predictors of students' retention [4]; Morrow and Ackermannexamine the relationship between students' intended retention and motivational attitude and sense of belonging [5]; Roberts and Styronexplore the relationship between student retention and their perceptions of social connectedness and satisfaction with faculty approachability [6].

Based on the above review of literature, it seems that the relationship between student retention and university selectivity is not fully studied. The following three research hypotheses are thus put forward to be tested in this study.

H1: University selectivity is positively correlated with student retention, controlling institution category (i.e., being public or private universities or colleges).

H2: Freshmen's ACT score mediates the relationship between University selectivity and student retention, controlling institution category.

H3: Average number of full-time faculty moderates the relationship between

university selectivity and student retention, controlling institution category.

Method

Data

To test the above-mentioned hypotheses, the National Center for Education Statistics was accessed and the IPEDS (Integrated Postsecondary Education Data System) Analytics: Delta Cost Project Database 1987-2010 in it was downloaded. Data fields extracted from the database are shown in Table 1. The selected data years are 2009 and 2010 academic years.

Table 1
Extracted data fields

Data Field	Data Year	Description
applicantcount	2009	Number of first-time degree/certificate-seeking applications received
admitcount	2009	Number of first-time degree/certificate-seeking students admitted
ft_faculty_per_100fte	2009	Full-time faculty per 100 FTE students
actcm25	2009	ACT Composite 25th percentile score
carnegie_sector_2005	2009	Carnegie Classification 2005 by Sector
ftretention_rate	2010	Full-time retention rate

Two filters were applied to the data set to limit the number of institutions that would be considered: (1) only four-year degree-granting institutions and (2) only institutions that reported all of the data fields in Table 1 were included. Applications of these two filters yielded a set of 328 institutions.

Variables

The dependent variable is student retention which is represented by ftretention_rate (i.e., full-time retention rate of 2010 academic year). The independent variables include: (1) university selectivity which is calculated by first, dividing admitcount (i.e., the number of first-time degree/certificate-seeking students admitted

of each institution in 2009 academic year) by applicantcount (i.e., the number of first-time degree/certificate-seeking applications received in the same year), and then use 1 minus the calculated score ; (2) Freshmen's ACT score which is represented by actcm25 (i.e., the ACT composite 25th percentile score reported by each institution in 2009 academic year); (3) Average number of full-time faculty which is represented by ft_faculty_per_100fte (i.e., full-time faculty per 100 FTE students reported by each institution in 2009 academic year); and (4) Institution category which dummy code institutions into 0 (public institution) and 1 (private institution) on the base of the Carnegie Classification 2005 by Sector. All variables are continuous except the control variable which is a categorical variable.

Analysis procedures

To test the three hypotheses, three procedures were conducted. For the first hypothesis, student retention was regressed on university selectivity, controlling institution category. For the second hypothesis, Baron and Kenny's approach was followed [7]. It involves the following three steps:

(1) Test whether the independent variable (university selectivity) predicts the mediator (freshmen's ACT score), controlling for institution category. If the effect is statistically significant, go to step 2.

(2) Test whether the independent variable predicts the dependent variable (student retention), controlling for institution category. If the effect is statistically significant, go to step 3.

(3) Test whether the independent variable and the mediator predict the dependent variable, controlling for institution category. Partial mediation occurs when the independent variable is a significant predictor in this step and full mediation occurs when it is not.

For the third hypothesis, first, independent variable (university selectivity)

and moderator (average number of full-time faculty), both continuous variables, were centered, then multiplied to get the cross-products. Next, a standard hierarchical multiple regression approach was used to test for moderation: control variable (institution category) was entered in the first model; centered independent variable and centered moderator in the second model; and then the interaction term (*i.e.,* the cross-products) in the third model.

Results

Descriptive Findings

Means, standard deviations, minimum and maximum for each variable are shown in Table 2.

Table 2
Statistics of variables

		private or public	university selectivity	Full-time faculty per 100 FTE students	ACT Composite 25th percentile score	Full-time retention rate of 2010 academic year
N	Valid	328	328	328	328	328
	Missing	0	0	0	0	0
Mean		.5518	.3493	5.9941	20.1738	.7578
Std. Deviation		.4980	.17544	2.67424	3.16184	.11061
Minimum		.00	.01	.98	13.00	.27
Maximum		1.00	.84	26.55	30.00	.99

Multiple Regression Testing H1

To test H1 (i.e., university selectivity is positively correlated with student retention, controlling institution category), student retention was regressed on university selectivity, controlling institution category. Then violations of assumptions were checked for.

Correlations. Correlations among the above-mentioned three variables (i.e. student retention, institution category and university selectivity) are shown in Table 3. The correlation between student retention and university selectivity is .324 (p<.001) and the correlation between university selectivity and institution category (public vs. private) is .146

(p<.01). As shown in Table 4, university selectivity and institution category together account for 11.0% of the variance in students' retentions. As shown in the ANOVA table (Table 5), the overall multiple regression is statistically significant (R^2=.110, $F_{[2, 325]}$=20.050, p<.001). H1 is supported. As shown in Table 6, the unstandardized regression coefficient (b) for university selectivity is .211 ($t_{[325]}$=6.317, p<.001) meaning that for each .01 points increase in university selectivity, students retention rate increased by .00211 points.

Table 3
Correlations

		Full-time retention rate of 2010 academic year	Private or public	university selectivity
Pearson Correlation	Full-time retention rate of 2010 academic year	1.000	-.023	.324
	private or public	-.023	1.000	.146
	University selectivity	.324	.146	1.000
Sig. (1-tailed)	Full-time retention rate of 2010 academic year		.338	.000
	private or public	.338		.004
	University selectivity	.000	.004	

Table 4
Model summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.331[a]	.110	.104	.10468

a. Predictors: (Constant), university_selectivity, private or public

b. Dependent Variable: Full-time retention rate of 2010 academic year

Table 5
ANOVA^b

	Model	Sum of Squares	df	Mean Square	F	Sig.
1	Regression	.439	2	.220	20.05	.000[a]
	Residual	3.561	325	.011		
	Total	4.000	327			

a. Predictors: (Constant), university_selectivity, private or public

b. Dependent Variable: Full-time retention rate of 2010 academic year

Table 6
Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.693	.014		50.399	.000
	private or public	-.016	.012	-.072	-1.359	.175
	University selectivity	.211	.033	.334	6.317	.000

a. Dependent Variable: Full-time retention rate of 2010 academic year

Multiple Regression Testing H2

To test H2 (i.e., freshmen's ACT score mediates the relationship between University selectivity and student retention, controlling for institution category), Baron and Kenny's three-step approach, as described in the above methodology part, was followed. Violations of assumptions were also checked for.

Mediation Results. First, M (mediator or ACT score) was regressed on X (influence or university selectivity) controlling for C (institution category). As Table 7 shows, the regression is statistically significant (Beta=.383, p<.001). Second, Y (Outcome or student retention) was regressed on X controlling for C. As Table 8 shows, the regression is statistically significant (Beta=.334, p<.001) which was also provided in the output associated with testing the first hypothesis. Third, Y was regressed on both X and M controlling for C. As Table 9 shown, the regression on the mediator is statistically significant (Beta=.772, p<.001) while the regression on the influence is not statistically significant (Beta=.038, p=.317). This suggests that freshmen's ACT composite 25 percentile score fully mediates the relationship between students' retention and the university selectivity while controlling for institution category. H2 is therefore supported. The following Sobel test shows that the indirect effect of university selectivity on student retention via freshmen's ACT composite 25 percentile score is statistically significant (see Figure 1: z=7.225, p<.0001).

Table 7
Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	19.61	.258		76.07	.000
	Control Variable	1.005	.347	.158	2.896	.004
2	(Constant)	17.40	.380		45.81	.000
	Control Variable	.651	.325	.103	2.006	.046
	Influence or IV	6.906	.921	.383	7.496	.000

a. Dependent Variable: Mediator

Table 8
Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.761	.009		83.27	.000
	Control Variable	-.005	.012	-.023	-.419	.675
2	(Constant)	.693	.014		50.39	.000
	Control Variable	-.016	.012	-.072	-1.359	.175
	Influence or IV	.211	.033	.334	6.317	.000

a. Dependent Variable: Outcome or DV

Table 9
Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.761	.009		83.279	.000
	Control Variable	-.005	.012	-.023	-.419	.675
2	(Constant)	.223	.025		8.906	.000
	Control Variable	-.034	.008	-.151	-4.256	.000
	Influence or IV	.024	.024	.038	1.003	.317
	Mediator	.027	.001	.772	20.169	.000

a. Dependent Variable: Outcome or DV

Input:		Test statistic:	Std. Error:	p-value:
a	6.906	Sobel test: 7.22492745	0.02580815	0
b	.027	Aroian test: 7.2203313	0.02582458	0
s_a	.921	Goodman test: 7.22963239	0.02579171	0
s_b	.001	Reset all	Calculate	

Fig.1. Sobel Test

Multiple Regression Testing H3

To test H3 (i.e., average number of full-time faculty moderates the relationship between university selectivity and student retention, controlling institution category), standard hierarchical multiple regression approach was used, as described above in the methodology. Violations of assumptions were checked later.

Moderation results. After entering control variable (institution category) in the first model, centered independent variable (university selectivity) and centered moderator (average number of full-time faculty) in the second model, and the interaction term in the third model, the regression results are shown in Tables 10 and 11. Centered independent variable and centered moderator accounted for 20% of the variance in student retention (F[2, 324]=40.504, p<.001). The addition of the cross-product only explained .6% of the variance in student retention, not a statistically significant increase (F[1, 323]=2.282, p=.132>.05).The interaction between university selectivity and average number of full-time faculty is not statistically significant. Thus H3 is not supported.

Table 10
Model summary

Model	R	R²	R² Change	F Change	df1	df2	Sig. F Change
1	.023ᵃ	.001	.001	.176	1	326	.675
2	.448ᵇ	.200	.200	40.504	2	324	.000
3	.454ᶜ	.206	.006	2.282	1	323	.132

Table 11
Coefficientsᵃ

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.761	.009		83.27	.000
	Control Variable	-.005	.012	-.023	-.419	.675
2	(Constant)	.773	.008		92.96	.000
	Control Variable	-.028	.011	-.127	-2.492	.013
	Xcentered	.146	.033	.231	4.363	.000
	MODcentered	.013	.002	.325	6.060	.000
3	(Constant)	.771	.008		91.15	.000
	Control Variable	-.028	.011	-.124	-2.433	.016
	Xcentered	.141	.033	.224	4.218	.000
	MODcentered	.011	.003	.272	4.224	.000
	XtimesMOD	.013	.009	.093	1.511	.132

a. Dependent Variable: Outcome or DV

To further understand the interaction between university selectivity and average number of full-time faculty, the latter variable was converted into a new, trichotomized variable based on the number of the full-time faculty per 100 FTE students. The regression of student retention on university selectivity for the three levels of number of full-time faculty per 100 FTE students (i.e, the lowest third, the middle third and the highest third of cases) were then plotted and three separate regression lines graphed. It appears that university selectivity is more positively correlated with student retention of institutions with high number of faculty per 100 FTE students than those with low or middle number (see Figure 2). Separate regressions find that the beta's for university selectivity to be similar between the low and the middle faculty number group while different from the high faculty number group: .122, .132 and .488 for low, middle, and high faculty number subgroups of colleges/universities (see Table 12, 13 and 14). In addition, the tables show that regression of student retention on university selectivity is not statistically significant for institutions with either the low (p= .205) or the middle (p= .170) number of full-time faculty per 100 FTE students (p=.205 for the former and p=.170 for the latter) while statistically significant for institutions with high number of full-time faculty per 100 FTE students (p<.001).
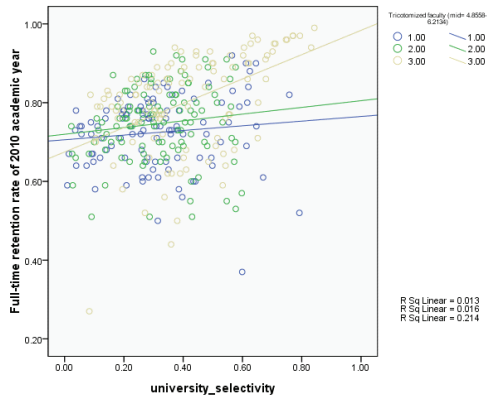
Fig. 2.Regression of student retention on university selectivity for three levels of faculty number.

Table 12
Coefficients[a,b]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.732	.011		66.830	.000
	Control Variable	-.022	.018	-.117	-1.214	.227
2	(Constant)	.735	.011		65.987	.000
	Control Variable	-.024	.018	-.125	-1.300	.196
	Xcentered	.064	.050	.122	1.275	.205

a. Dependent Variable: Outcome or DV
b. Selecting only cases for which Tricotomized faculty (mid= 4.8558-6.2134) = 1.00

Table 13
Coefficients[a,b]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.759	.014		56.088	.000
	Control Variable	-.022	.018	-.115	-1.202	.232
2	(Constant)	.762	.014		55.742	.000
	Control Variable	-.023	.018	-.122	-1.278	.204
	Xcentered	.090	.065	.132	1.383	.170

a. Dependent Variable: Outcome or DV
b. Selecting only cases for which Tricotomized faculty (mid= 4.8558-6.2134) = 2.00

Table 14
Coefficients[a,b]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.833	.024		34.803	.000
	Control Variable	-.041	.028	-.143	-1.495	.138
2	(Constant)	.826	.021		39.285	.000
	Control Variable	-.059	.025	-.203	-2.408	.018
	Xcentered	.326	.056	.488	5.769	.000

a. Dependent Variable: Outcome or DV
b. Selecting only cases for which Tricotomized faculty (mid= 4.8558-6.2134) = 3.00

Discussion

From the above analysis, three conclusions can be made. First, university selectivity is positively correlated with student retention. The more selective a university is when recruiting students, the higher student retention it has. Second, college freshmen's ACT scores fully mediate the relationship between university selectivity and student retention. This suggests that the relationship between university selective and student retention is basically indirect via freshmen's ACT scores. The third conclusion is, contrary to proposed hypothesis (H3), average number of full-time faculty does not moderate the relationship between university selectivity and student retention.

This study suggests that student retention is influenced by university selectivity and student's performance in high school. It is also indicated that one of the important reasons students drop out is poor academic performance or unpreparedness for academic study. Interventions of particular programs at student's different critical stages of their learning journey should be helpful to reduce students' dropout rate. On the other hand, student's high school academic achievement still should be one of the most important criteria in admission decision of college freshmen.

References

R. Chen, "Institutional characteristics and college student dropout risks: A multilevel event history analysis," Research in *Higher Education, vol. 53*(5), 2012, pp. 487-505.

D. Lee, E. A. Olson, B. Locke, S.T. Michelson, andE. Odes, "The effects of college counseling services on academic performance and retention," *Journal of College Student Development, vol. 50*(3), 2009, pp. 305-319.

T.C. Lin, W.W. C. Yu, and Y.C. Chen, "Determinants and probability prediction of college student retention: new evidence from the Probit model," *International Journal of Education Economics and Development, vol. 3*(3), 2012, pp. 217-236.

J. A. Morrow, and M. E. Ackermann, "Intention to persist and retention of first-year students: *The importance of motivation and sense of belonging,"* College Student Journal, vol. 4*6(3), 2012, pp. 483-491.

J. Roberts, and R. Styron, "Student satisfaction and persistence: Factors vital to student retention," *Research in Higher Education Journal, vol. 6*(3), 2010, pp. 1-18.

R. M. Baron, andD. A. Kenny, "The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations," *Journal of Personality and Social Psychology, vol. 51*, 1986, pp. 1173-1182.