

College Rankings: Creating an Equitable Model of Transformation and Institutional Effectiveness

Allison L. Hurst, Oregon State University

Abstract

Colleges that serve working-class students show up poorly in traditional rankings of US colleges. Without appropriate outcome measures, measures of ‘quality’ of inputs drive most current ranking systems. The trouble is that quality is often just a measure of pre-existing privilege (e.g., selectivity, average SAT scores). In this article, I demonstrate the viability of a model that uses economic returns data while factoring in the relative lack of privilege of students attending any particular institution as a way of ranking that institution’s transformative efficacy and institutional effectiveness. The model was then tested on a diverse sample of 655 US colleges and universities for whom reliable economic returns and institutional effectiveness data are available. Unlike widely used rankings models, this proposed alternative model can distinguish between reproducing privilege (high economic returns as expected, low defaults, timely year to degree and fewer incompleters) and facilitating social mobility (higher returns and persistence than would be expected given the incoming characteristics of students). The article concludes with a discussion of the uses to which such a model could be best put.

Keywords

College rankings, working-class college students, social mobility, institutional effectiveness

Introduction: what do college rankings measure – prestige or outcomes?

The most widely known, used, and criticized college rankings model belongs to *U.S. News & World Report*. USNAWR first published its annual ‘America’s Best Colleges’ rankings in 1983. A key part of the rankings are based on reviews by peer institutions. This has made it difficult for colleges with strong academic reputations to be displaced in the rankings. Indeed, a handful of colleges and universities have remained at the top of the rankings since the first publication. Other criteria used to rank colleges are its rate (the percentage of admitted students), average SAT scores of admitted students, and yield rates. All three of these criteria reward colleges for their being choosy about whom they admit. For this reason, open admissions colleges or colleges that serve populations who may be underprepared for college never make it into the top of chart.

Sociologists, economists, and educational researchers have repeatedly demonstrated the weaknesses of this hegemonic ranking system (Ehrenberg 2003, Brooks 2005, Leo 2011, Volkwein & Grunig 2005). It is a particularly poor measure of institutional effectiveness, as ‘numerous factors that frequently lie beyond institutional control strongly influence degree completion rates, such as the socioeconomic status and academic preparedness of incoming

students' (Horn and Lee 2016 pg.470). While USNAWR has long been criticized, it's also been viewed as the only game in town. More than two million viewed its 'Best Colleges' webpage in 2014, according to USNAWR, and most parents and potential college students have probably allowed its rankings to factor into their decision about what college to attend. In fact, moving into its 'Top 50' substantially improves admission (number and quality of admissions) the following year (Bowman & Bastedo 2099). Colleges thus feel bound to make decisions based on how these decisions will affect their ranking. Attempts to displace these measures have failed in the past. In 2015, for example, the Obama administration retreated from a plan to rate colleges based on completion rates and time to degree, two measures that might have shifted focus to measuring student success. It is worth noting, however, that such a plan might also have effectively punished schools that accept large numbers of underrepresented students, as these students often take longer to complete a degree (Field 2013).

One of the largest problems with all existing and planned rankings systems is the absence of an agreed upon measurement outcome. USNAWR's rankings rely heavily on the quality of *inputs* (e.g., selectivity measures). In other words, a school's ranking is determined less by student success *through and after* college and more on the *quality of students admitted*. The rejected Obama plan considered a variety of outcome measures – completion rates, transfer rates, employment and earnings, graduate school attendance, and student loan repayment rates. No agreements were ever made on how these measures would be used, in what proportions, or, as with the case of employment and earnings, how obtained. Ideally, we would like an outcome measure that captures learning and growth over time (e.g., Arum and Roksa 2010), but graduates and their parents might prefer a more basic measurement of economic payoff – do graduates find jobs that pay well? Working-class college students in particular need to know how choosing college X over college Z may affect one's economic security ten years down the line. This is basic consumer information nearly impossible to acquire in the field of higher education and its absence has perhaps allowed college rankings to become the game of prestige that they presently are.

No federal or state agency regularly collects employment and salary information on college graduates of particular institutions. This is the reason we have so far been unable to rate individual colleges and universities using an outcome measure of performance. Into this vacuum has arrived Payscale, a for-profit company that has used crowdsourcing to amass a huge amount of data on the beginning and midcareer salaries of approximately 1000 colleges and universities. To date, more than thirty-five million profiles have been added to their database. The company sells individual reports on companies and colleges to individuals and investors but makes the aggregate median average pay available to all. Reported midcareer salaries on Payscale range from \$40,300 (Shaw University) to \$138,800 (Harvey Mudd College), with 50% falling between \$65,000 and \$85,000. With this tool, we can devise an alternative rankings system that actually measures economic returns. There will still be a place, of course, for rankings of academic prestige. If, however, the administration wants to reward schools that make a difference in the lives of students, it will need a model very different from the USNAWR's 'Best Colleges.'

In this article, I demonstrate the viability of a model that uses economic returns data while factoring in the relative lack of privilege of students attending any particular institution as a way of ranking that institution's transformative efficacy. I offer this as a heuristic model, and leave mathematical refinements and implementations to other researchers. With publicly available data, any person

can use these models to assess an institution's overall effectiveness and thereby compare institutions. I offer suggestions for such comparisons for legislators and institutions later in the article.

Measuring outcomes in a stratified higher education field

The higher education system in the US is less of a system than a configuration of often autonomous regional and institutional networks. Unlike many European systems, where students compete to find a place in a fixed hierarchical system, such as entry to the *grands écoles* vs. universities in France, students in the US confront a bewildering set of options. They can choose to enter a two-year program or a four-year program, public or private, for-profit or non-profit, comprehensive or liberal arts or technical, religious or non-denominational, in-state or out-of-state, to name only the most obvious choices. Surely part of the appeal of the *US News system* and rankings in general is to provide some clarity to this universe.

Although we often talk about having a meritocracy in the US, the complexity of the higher education system makes it difficult to properly assess this claim. Some schools cost much more than others, and they may or may not be any more 'selective' than less expensive schools. Some of the most academically rigorous institutions are state flagships (for example, UC Berkeley) that have low in-state tuition and so should be open to the best and brightest from all social backgrounds. On the other hand, some private universities are quite expensive, have long histories of serving elite students, and are also incredibly selective in their admissions policies. The recent cheating scandal involving celebrities in the US buying access to elite universities is a case in point. An excellent study by Winston and Hill (2005) concludes that many smart working-class students never apply to these institutions. We have a system that, perhaps partly due to its bewildering complexity, often sorts students into particular parts of the system based on factors that have less to do with ability and more on perceived affinities and expected costs.

Researchers have long remarked upon the ways that class background intersects with the US higher education landscape (Berg 2010; Crane 1969; Engberg 2012; Grodsky & Jackson 2009; Kahlenberg 2004; Reynolds 1927; Stuber 2011). Beyond the 'public vs. private' or 'four-year and two-year' divide, there are many ecological niches. In general, low-income and working-class students are less likely to attend college and more likely to attend a less selective institution when they do (Walpole 2007). Astin and Oseguera (2004) shockingly reported that US higher education was more socioeconomically stratified at the time of their study than at any time during the past three decades, with elite students often crowding out middle-income students at the top colleges. Things do not appear to have reversed course. Low-income, working-class, and first-generation students are particularly underrepresented at selective private liberal arts colleges, and their difficulties attending such colleges have been well documented (Ostrove & Long 2007; Aries & Seider 2005; Mullen 2009).

Colleges and universities in the US are also raced and gendered. There are currently forty-two four-year institutions serving women exclusively, and sixty-six serving men exclusively (NCES *College Navigator* data, author analyses). Ninety-one Historically Black Colleges and Universities (HBCU) and twelve Tribal Universities currently operate in the US and there are many more newly emerging (newly identified, at least) Minority-Serving Institutions (defined as enrolling 25% or

more of a particular ethnic group). What these latter designations are capturing are historic ties between particular demographic groups of students and particular institutions, but these exist to some extent at all colleges and universities. Ask an ‘enrollment manager’ and they will be able to tell you exactly what the ‘target’ student is for a given campus. The universe of Minority-Serving Institutions (MSI) is diverse in terms of class. Some colleges, like Howard University, a private HBU, are difficult to get into, expensive, and serve many elite students while others, like South Carolina State University, a public HBU, are relatively easy to get into, affordable, and serve mostly low-income students (NCES *College Navigator* data, author analyses).

The complex stratification of the American higher educational universe makes measuring institutional effectiveness extremely difficult. A recurring question in the literature is how much of a particular outcome (whether it be learning, retention rates, employment, income, and even satisfaction) can be attributable to the institution and how much is more of a product of the types of students the institution serves (Alon 2007; Hussey & Swinton 2011; Kim, Kim, Jacquette & Bastedo 2014; Mattern, Shaw and Kobrin 2010; Melguizo 2010; Sibulkin 2011). Studies that show different economic returns by *type* of college attended often highlight the difficulties of evaluating institutional type effects when types serve and attract significantly differentiated student bodies. For example, Thomas and Zhang (2005) demonstrate that graduates from what they refer to as ‘higher quality’ colleges earn more early in their career, but leave one wondering, ‘do elite students earn more because they attend elite colleges, or because they are elite’ We still don’t really know (Brand & Halaby 2006). We do know that returns to attending highly selective colleges have been increasing since 1972 (Hoxby 2001), but how much this is attributable to greater horizontal stratification within higher education is not known.

Students are increasingly concerned with the costs of college. An all-time high of 88% of first-year students in 2011 reported that they went to college ‘to be able to get a better job’ (Pryor et al. 2012). A majority (67%) believed that the current economic situation affected where they enrolled. This may amplify stratification within higher education. It also puts a great deal of pressure on public colleges and universities, whose students tend to be most concerned about post-graduate outcomes, to demonstrate their institutional effectiveness. Many state legislators are expressing concern about the rising costs of college and demanding results. Some states, like Maine, are linking unemployment insurance wage records with college student records to measure the impact of various colleges and major fields of study (Leparulo 2015). There are privacy concerns with this type of analysis, however, and most states currently block use of this data.

Researchers have explored the impact of college on the economic outcomes of graduates, and the differences of this impact for various groups and majors (Brand & Yu 2010; Hu & Wolniak 2010; Liu, Thomas & Zhang 2010; Torche 2011). Put simply, college does not seem to have the same returns for all students, although the patterns are quite complicated as different majors and different types of institutions interact with student background to produce a wide array of possible outcomes. For example, attending an elite institution seems to make a bigger overall impact on low-income and racial minority students, although these students still earn less than their peers (Brand & Yu 2010). Female graduates appear to earn less than their male peers, even controlling for differences in fields of study (Zhang 2008). Payoffs to more ‘practical’ majors also may appear larger immediately out of college than in the long term (Roksa & Levey 2010), but this may have less to do with the major and more with the types of students attracted to practical vs. liberal arts

majors (Goyette & Mullen 2006). Suffice it to say, research into the forms and mechanisms of ‘horizontal stratification’ within higher education is an ongoing project (Garber & Cheung 2008). As useful as this line of research has been, it has not linked outcomes to particular institutions. Research at the institutional level often lacks economic returns data. For example, Eff (2012) evaluated the cost effectiveness of roughly 1000 US four-year colleges and universities. He generated scores that allowed the ranking of institutions by a weighted sum of institutional characteristics per dollar of average net price, arguably providing an objective means of ranking institutions as the best values in higher education. But Eff’s ranking does not include any measure of economic returns to the student. The rankings tell us how effective a college is in the use of its dollars, but not how this effectiveness translates to better outcomes for its graduates. What we need is a similar measure of individual institutional effectiveness of outcomes, taking into account the types of students that attend the institution.

Constructing a transformative effectiveness ratings system for colleges and universities

This section will describe the methods and criteria used to generate a model that allows comparisons of transformative effectiveness by institution. It is important to first be very clear about the purpose of the model in order to include criteria that advance that purpose. What will from now on be referred to as the Undergraduate Transformative Effectiveness Ratings Model (UTERM) provides comparative measures of average outcomes of students of a given college or university in terms of midcareer salaries *taking into account* the school’s overall demographic profile. Its purpose is to evaluate how well the graduates of a particular institution fare economically *given their averaged points of origin*. Unlike *US News* ratings, colleges and universities are not penalized for having working-class students; in fact, institutions that only serve elite students will NOT score well in UTERM, because they will be unable to show transformative effects. This means, too, that any attempts to ‘game’ the system by increasing selectivity will have no or negative effects. Conceptually, there are only two ways to advance one’s place in the UTERM ratings – either increase participation of underserved students or improve economic returns for graduates.

The primary users of such a model would be those interested in comparing institutional effectiveness for research or policy purposes. State legislators, for example, might use the model to compare effectiveness of its public institutions, or to see how the flagship university holds up against competitor private institutions in the state. Researchers might be drawn to examining how types of institutions (HBCUs, liberal arts colleges in the Northeast) are more or less transformative. The model is not intended for individual student applicants, for each individual’s outcome may differ substantially from the averaged transformative effect. Because of the primary use for policy and research, the following criteria were included: (a) the model should be based on publicly available data; (b) the model should not rely on proprietary information or formulae that are not publicly made available; (c) the model should be easy to adjust (i.e., shifting the relative weight of a particular factor for purposes specific to the observer); (d) the model should include both an overall measure of institutional effectiveness as well as different benchmarks for different types of institutions (given the variability of demographic profiles by institutional type).

The model was created in stages, testing for validity and reliability with the successive addition of cases. The possible universe was limited to 1000, as that is the total number of colleges for which

Payscale offers reliable midcareer salary data. The sample of 655 was chosen by pulling an equal number of colleges from each salary point, while purposively including the flagship university of each state and all AAU member institutions. Forty-eight HBCU are included in the sample. For profits were originally included (as they are included in Payscale's database) but were found to not work well with the model; this could be that the information they provide is not as complete or reliable. Future research will need to be done in this area.

The first step in creating UTERM was an examination of the relationship of key demographic data with economic returns. Relationship to returns scatterplots were run for (a) percentage of underrepresented minority students; (b) percentage female; (c) percentage receiving Pell Grants (federally provided need-based grants that are often used in the US as an indicator of low-income status) relative to average midcareer salaries. Also examined was the relationship between cost of college and these variables and cost of college and returns. Linear regression models examined selectivity and SAT scores' relationship to economic returns and the demographic factors listed above. The results of many of these preliminary tests will be summarized below.

Only after examining the relationships between potential variables of interest was the UTERM model created. UTERM takes into account the relative lack of privilege of the admitted students, as described in more detail below. Colleges with good economic returns will score higher if their 'inputs' were lower than if they were higher. For example, a college whose graduates earn the same amount as another college but whose students scored 50% lower on the SATs will rank higher in the model. Colleges with the highest inputs should have much higher economic returns; if they are merely the same or slightly greater those colleges will not earn a very high score (although they may and probably will earn a good score). This is why the model is said to measure *transformation*. Schools with high scores are those that provide a decent economic return for a group of students whose future prospects may not have been assured at the outset. In other words, schools whose students do better than predicted, given where they began (e.g., low-income students, underrepresented students, non-traditional students, students with low SAT scores). In contrast, schools with low scores are those that fail to provide a decent economic return at all, for any group of students, but particularly for those who entered college with high SATs and high expectations.

UTERM is designed to be simple to use and share. Because we are not trying to measure academic quality or prestige, or proxies therefore, we can rely on a few key points of information, obtainable through the National Center for Education Statistics' *College Navigator* program, coupled with the information provided by Payscale.

There are three parts to the construction of a score. First, our outcome variable, *Economic Returns*, is constructed by factoring the reported annual net cost of tuition by the reported number of years to graduate (ranging from 4 to 6) and subtracting this from a constructed variable of five years' worth of midcareer salary as reported by Payscale.

Second, I constructed a 'lack of privilege' score, a variable summarizing the key demographic information examined above. This score was tested using linear regression and found to negatively predict economic returns (described in greater detail below in results). The colleges with the greatest residual scores do show up as particularly high (or low) in our rankings scheme. In other

words, we can predict fairly well what the economic returns are of a particular college given its demographic makeup. This should not be a surprise to sociologists of education and higher education, but it is worth stressing here for the reader unfamiliar with this area of research. Schools that primarily serve working-class students generally have graduates who make less money than graduates of other institutions. This is true even for academically rigorous colleges whose working-class students scored high on the SAT. In contrast, the more elite the body of the school, the more college ‘pays off.’

There are four elements to the *Lack of Privilege* variable: percentage of admitted students receiving Pell Grants, percentage of Black, Latino/a, and Native American students, percentage female, openness of admissions (here selectivity works against colleges not in their favor), and distance of average SAT scores from a perfect score. In our sample, Savannah State University has the highest constructed *LOP* score, 3.14. Eighty-percent of Savannah State’s students receive Pell Grants, 89% are underrepresented minorities, the school admits 83% of its applicants, fifty-five percent of its students are female, and the average SAT score (math and reading) was 845 out of 1600. On the other end, Washington University in St. Louis (closely followed by University of Chicago and Harvard University) had the lowest *LOP* score. Only 5% of Wash U’s student received Pell Grants, 11% were underrepresented minorities, 51% were female, only 16% of applicants were admitted, and the average SAT score was 1485 (out of 1600). The final score for Wash U was 0.610.

From these two variables I created an unadjusted score. For the two examples above, we note that Savannah State has a predictably low midcareer salary, \$51,500, and Wash U a predictably high one of \$107,000. At the same time, it cost \$55,000 to earn a degree from Savannah State (even with an average time to graduate of 5.71 years) while it cost \$136,767 to graduate from Wash U (4.15 years). Unadjusted, Savannah State scores 85 and Wash U scores just 32. We then make two adjustments. We note that Savannah State has a very high dropout rate, in addition to the almost six years’ time to graduate. In fact, 66% of its students never graduate. It also has a very high loan default rate – 21%. In contrast, 94% of Wash U’s students graduate and only 2% default on their loans. Adjusting for these ‘bad outcomes’, Savannah State’s final adjusted score is just 43.96. This puts it in the bottom quarter of the sample. Wash U, meanwhile, falls to 28.59. It is in the bottom decile of the sample. Neither of these two colleges are very transformative. In Wash U’s case, this is because its relatively elite students do slightly worse than expected, once the high cost of attendance is factored in. In Savannah State’s case, it is because it is failing a large proportion of its admittedly challenging students. In contrast, South Carolina State University has a very similar *LOP* profile but has higher economic returns and fewer defaulting students and scores in the top quartile.

Limitations

Perhaps the most controversial element of UTERM is the reliance on midcareer salaries as its sole outcome measure. UTERM measures economic returns. It is neither a measurement of academic prestige nor excellence. Indeed, there is nothing at all about UTERM that evaluates the quality of instruction, instructors, campus climate, student subjectivity, or college infrastructure. UTERM is not a measure of learning-based institutional effectiveness. There are many things that make going to college worthwhile other than the projected midcareer salary upon graduation. Many readers will find the lack of other outcome measures a limitation, but this was a methodological choice.

UTERM attempts to address the question, and only this question, of ‘do graduates of X college have better or worse economic returns than would be predicted based on their incoming attributes?’ The data on midcareer salaries relies on aggregated self-reported data of graduates, as provided by Payscale. As described earlier, Payscale is a for-profit company that has used crowdsourcing to amass a huge amount of data on the beginning and midcareer salaries of approximately thirty-five million graduates of more than 1000 colleges and universities. Because graduates without jobs or graduates with poor jobs are less likely to add their data, the returns are probably higher than they are in reality. That said, the inflated returns should hold constant across the database. The lack of reliable official data on economic returns of college graduates linked to institutions is a limitation of this study and any other studies exploring the value of higher education today. This is particularly noteworthy regarding for profit institutions, which were not included here for lack of reliable returns information.

The undergraduate transformative effectiveness ratings model (UTERM)

Table 1 provides descriptive data of the sample for some of the key variables used in the analysis. The average SAT score (out of 1600) is about 100 points higher than the national average, perhaps reflecting the exclusion of for-profits from the sample and the purposive inclusion of all AAU institutions. Note the range of reported midcareer salaries from Payscale are generally in line with what we know from Census data. Although there are probably selection biases involved in the self-reporting of salaries, the range of such salaries makes this a reasonably reliable measure of the aggregate economic returns of particular institutions.

The categories of region and type are mutually exclusive. Colleges and universities (C&U) were coded as ‘elite’ if they belong to the Ivy League, ‘liberal arts’ if they adhere to the liberal arts model and are not already coded as ‘elite’, ‘private’ if neither state-supported nor operating under a liberal arts model, ‘flagship’ if they are one of the flagship public universities of their state and ‘public’ if a public four-year not a flagship. Two other categories, HBCU and Specialty (e.g., Cooper Union, West Point, California Institute of the Arts) are not shown here but are included in the overall sample. There is clear demographic variation by region and type. For example, forty-one percent of students attending non-flagship universities receive Pell Grant assistance compared to less than a quarter at flagship universities. The proportion of Pell Grant recipients is highest among C&U in the South (42%). Similarly, the percentage of underrepresented minority students varies greatly by region (38% in the South compared to 13% in the Midwest). Table 1 also shows that years to graduate, percentage of incompleters (those beginning but failing to graduate), and default rate vary substantially by type of college and region. Southern C&U Have long years to graduate, high numbers of incompleters, and a very high cohort default rate. Years to graduate are longest at public universities. Flagship universities have a lower percentage of incompleters and a lower cohort default rate than any other type of C&U.

Table 1. Means of UTERM sample, by Region and Type of Institution

	Total	REGION				TYPE ¹			
		Northeast	Midwest	South	West	Private	LibArts	Flagship	Public
%Female	0.55	0.56	0.53	0.57	0.55	0.56	0.58	0.53	0.55
%Pell Grant	0.35	0.28	0.31	0.42	.036	0.30	0.30	0.24	0.41
%BLAI	0.25	0.18	0.13	0.38	0.23	0.20	0.16	0.14	0.27

% Admit	0.61	0.55	0.66	0.59	0.65	0.61	0.59	0.65	0.68
SAT	1096	1153	1116	1048	1088	1119	1140	1176	1043
Net Price (in US dollars)	20,669	23,418	20,726	18,514	18,532	26,432	25,220	14,722	13,140
Midcareer Salary (in US dollars)	82,105	85,388	80,138	77,530	83,761	83,425	80,167	83,190	77,117
YTG	4.80	4.52	4.68	4.95	4.96	4.53	4.37	4.85	5.24
Incompleters	0.39	0.30	0.35	0.46	0.40	0.35	0.32	0.29	0.47
Default rate	7.90	5.38	6.13	11.22	6.94	6.03	5.86	5.60	9.25
<i>N</i>	655	162	134	231	128	169	152	50	195

(1) Types included in the total sample but not reported here are Elite ($n=9$), HBCU ($n=48$), and specialty ($n=30$).

The relationship between the demographic profile of a college and negative outcomes can be visually plotted. Figures 1-3 demonstrate the relationship between three key pieces of demographic information (percentage of Pell Grant recipients, percentage of underrepresented minorities, and SAT scores) and incompleteness, extended graduation rate, and cohort defaults. The percentage of incompleters clearly correlates with both low-income status and low SAT scores, and somewhat with the racial composition of a C&U. Years to graduate shows a more tenuous, but still apparent, connection, between these three demographic variables. Defaults are clustered among C&Us that serve low-income and minority youth. They also reduce as SAT scores increase

Figure 1. Incompleters by (a) Low-income students; (b) BLAI students; and (c) SAT scores

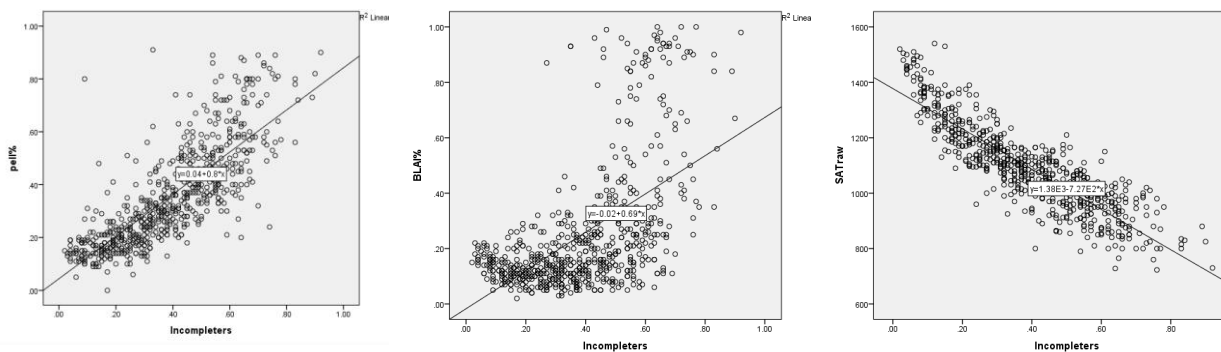


Figure 2. Years to Graduate by (a) low-income students; (b) BLAI students; and (c) SAT scores

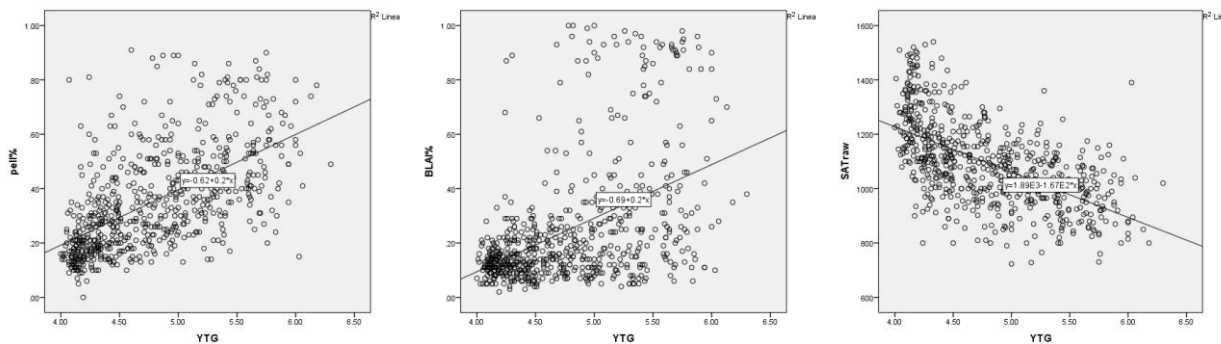
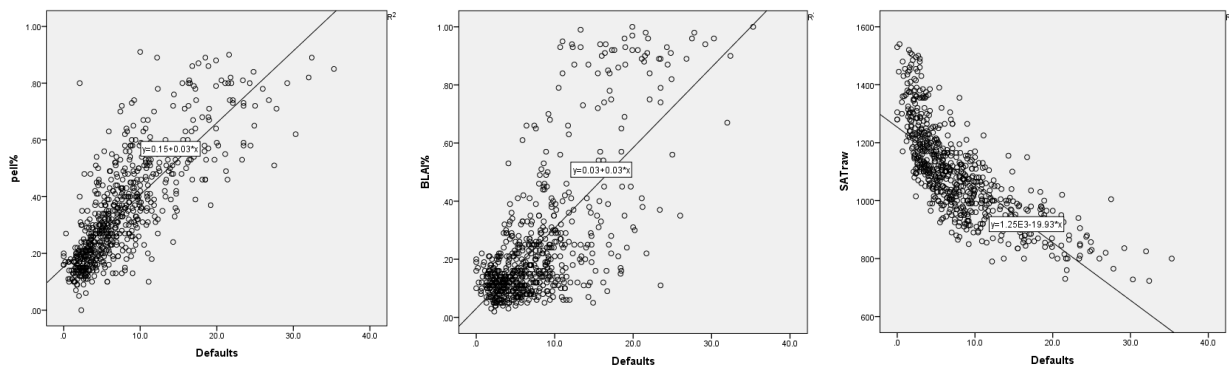


Figure 3. Defaults by (a) low-income students; (b) BLAI students; and (c) SAT scores



Regression models were run for each of these particular outcomes, first running the entire demographic profile (including percent female) and secondly with added controls (selectivity, net price, region, and type). The correlation between percentage low-income (Pell Grant recipients) was statistically significant at the $p < .001$ level for incompleters (Table 2). Racial composition was highly statistically correlated with years to graduation, even more so with added controls. Once controls of region, type, price were added, low-income status was not a significant predictor of years to graduate (Table 3). Both income and race were very significantly correlated with default rates (Table 4). Similar relationships were found correlating race, class, gender, and SAT scores with midcareer salaries (regression results available upon request).

Table 2. Percent Incompleters Regressed on Student Background Characteristics (Equation 1) and School Conditions (Equation 2)

Variable	Equation 1		Equation 2	
	<i>b</i>	<i>Beta</i>	<i>b</i>	<i>Beta</i>
% female	-.087* (.033)	-.048	-.075* (.034)	-.041
% low-income	.299*** (.036)	.300	.259*** (.037)	.260
% BLAI	-.018 (.022)	-.022	.011 (.024)	.013
SAT	-.001*** (.000)	-.656	-.001*** (.000)	-.593
% admit			.057* (.023)	.061
Net price			-.019*** (.004)	-.112
Region			.007* (.003)	.040
Type			-.002 (.003)	-.020
Constant	1.196		1.117	
R ²	.790		.804	

Table 3. Years to Graduation (of those completing) Regressed on Student Background Characteristics (Equation 1) and School Conditions (Equation 2)

<i>Variable</i>	Equation 1		Equation 2	
	<i>b</i>	<i>Beta</i>	<i>b</i>	<i>Beta</i>
% female	-.631*** (.165)	-.120	-.223 (.135)	-.043
% low-income	.806*** (.177)	.279	.174 (.146)	.060
%BLAI	.233* (.107)	.099	.333** (.095)	.141
SAT	-.001*** (.000)	-.326	-.001*** (.000)	-.190
% admit			.286*** (.089)	.106
Net price			-.219*** (.016)	-.448
Region			.033* (.013)	.065
Type			.043*** (.011)	.124
Constant	6.040		5.619	
R ²	.394		.630	

Table 4. Percent Defaulters Regressed on Student Background Characteristics (Equation 1) and School Conditions (Equation 2)

<i>Variable</i>	Equation 1		Equation 2	
	<i>b</i>	<i>Beta</i>	<i>b</i>	<i>Beta</i>
% female	-6.89*** (1.21)	-.123	-5.80*** (1.26)	-.103
% low-income	9.48*** (1.30)	.305	9.35*** (1.37)	.301
%BLAI	7.99*** (0.79)	.313	7.30*** (0.90)	.286
SAT	-.01*** (.00)	-.352	-0.01*** (0.00)	-.351
% admit			-0.21 (0.83)	-.007
Net price			0.09 (0.15)	.016
Region			0.10 (0.12)	.018
Type			0.28** (0.10)	.075
Constant	20.80		18.95	
R ²	.717		.721	

Although the relationships among key variables were empirically demonstrated, the UTERM model was developed theoretically, as a way of measuring the transformative effects of particular colleges and universities. The variables were chosen based on previous research in the sociology of education and social reproduction theory generally. There was no attempt to determine exactly

how much race, class, or gender differences accounted for differing economic returns. As described above, a ‘lack of privilege’ (LOP) index was constructed using the percentage of Pell Grant recipients, underrepresented minorities (Black, Latino/a, Native American and Pacific Islanders), female students, distance of average SAT from perfect score, and percent admitted. The percentage of female students was included in the LOP index because women still earn less than men when graduating from college (something noted empirically when women’s colleges were scoring lower than otherwise expected in an early model). The LOP index was created by summing all four elements, although percent admitted was discounted by half. This is a relatively easy way for those interested to compare the privilege profile of colleges and universities of interest to them. Table 5 reports the mean average LOP scores by type of college. Unsurprisingly, elite colleges show very low LOP scores while public universities and HBCUs report high scores. Other researchers may wish to alter the relative weights of the four elements and, again, this is easy to do. For example, reducing the weight of the underrepresented minority variable would significantly alter the LOP score of HBCUs.

Figure 4. Adjusted Scores plotted against Raw Scores for Elite Colleges

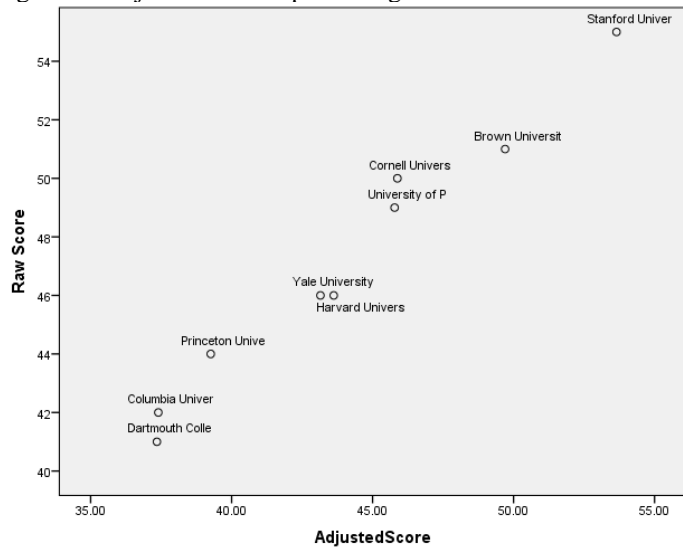


Figure 5. Adjusted Scores plotted against Raw Scores for Private C&U in South

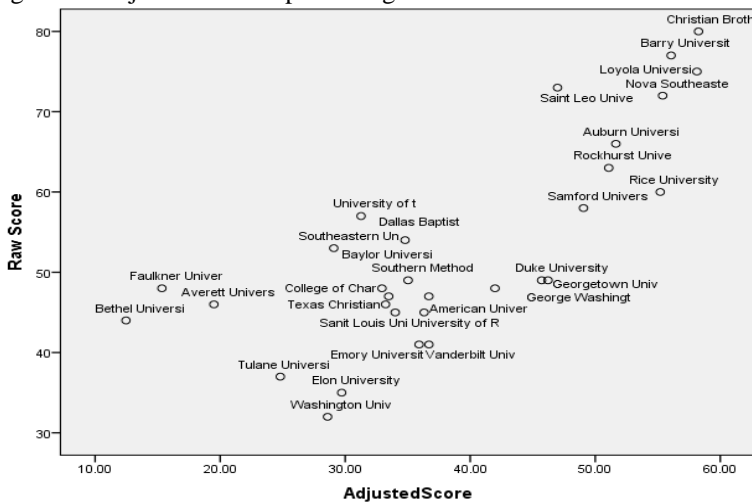


Table 5. Means of UTERM constructed variables, by Type of Institution

	Total	Elite	LibArts	Private	Flagship	Public	HBCU
LOP score	1.55	0.74	1.42	1.50	1.39	1.82	2.47
Total Cost of College (in US Dollars)	94,954	86,764	108,214	117,887	71,118	67,755	112,599
PayOut (in US Dollars)	631,319	956,250	585,238	598,478	689,660	635,655	484,469
RawScore	62.43	47.11	53.05	57.01	63.25	76.34	78.67
AdjustedScore	47.66	43.98	41.45	44.17	50.14	56.47	47.14
<i>N</i>	655	9	152	169	50	195	48

Table 5 also includes mean averages for the other constructed variables used in UTERM, as well as raw and adjusted scores. Total cost of college was constructed by multiplying the net price (as provided by NCES data) by the average years it takes to graduate from the school. As public universities tend to have longer years to graduate, this undercuts some of the cost differential between these institutions and more expensive private institutions. Even so, public C&U have the lowest total cost while private the highest. Both flagship universities and elite C&U have lower than average total costs. PayOut is measured by taking a ten-year multiple of the average midcareer salary provided by Payscale and deducting the total cost of college. Elite C&U clearly outperform in this area. Liberal arts colleges and HBCUs (perhaps attributable to high total costs) do the least well in terms of economic rewards. The RawScore simply factors PayOut by LOP. C&U with LOP scores lower than ‘1’ will see their economic returns discounted, while schools with high LOP scores will see their economic returns amplified. Perhaps not surprisingly, public universities have higher raw scores than private universities. Elite universities show the lowest. There is a great deal of variation at the institution-level, however, as will be described further below.

Raw scores can be useful indicators of a particular institution’s transformative effects, but they do not perhaps differentiate sufficiently between similarly situated schools. The adjusted score modifies the raw score by factoring in defaults and incompleters (note that years to graduate was already factored in as part of the total cost of college). Various factors and strengths were modeled to find a formula that could differentiate between similar institutions without undercutting the main strength of UTERM. Deducting Percent Incompleters*20 and Cohort Default Rate*13 from the Raw Score was found to be optimal. Table 5 reports the adjusted scores by type of institution. The larger variation by type has been reduced but still exists. At the same time, variations at the institution level are more apparent. The formula can be rendered as:

$$\text{LOP score} * \text{Economic Returns} = \text{Raw Score} * \text{Bad Outcomes score} = \text{Adjusted Score}$$

Figures 4 and 5 demonstrate the effectiveness of the adjustment. When adjusted scores are plotted against raw scores for elite colleges, there is very little adjustment taking place (Figure 4). Elite schools have low default scores and very few incompleters. Brown University shows less impairment relative to the other schools in the group. In contrast, private colleges and university in the South show a great deal of institutional variation in adjusted scores relative to raw scores

(Figure 5). Faulkner University, for example, has low adjusted core, relative to its raw score, indicating its students are dropping out and defaulting more than would be expected, while Georgetown and Rice University are showing less negative outcomes than their place in the field would predict.

Table 6. Raw and Adjusted Scores of Selected Institutions

Institution	Raw Score	Adjusted Score	Type of Institution
UT-El Paso	148	122.31	Public
Howard University	108	87.11	HBCU
Brandeis University	76	70.98	Private
UC-Berkeley	66	59.65	Flagship
Brown University	51	49.7	Elite
Hiram College	68	49.2	Liberal Arts
Reed College	57	48.07	Liberal Arts
Duke University	49	46.22	Private
U Alabama	63	44.7	Flagship
Ithaca College	47	37.62	Liberal Arts
Western Oregon U	52	31.36	Public
Morehouse College	70	27.5	HBCU
Guilford College	33	12.19	Liberal Arts

Table 6 provides raw and adjusted scores for a variety of types of institutions. Although it is difficult for elite, private, and liberal arts colleges to score very high in UTERM, there is still significant variation within types as this table indicates. Complete results are available upon request.

A new typology: reproducers of privilege, transformative colleges, safe bets, and troubled returns

One of the major advantages of UTERM is the ability to distinguish colleges that merely reproduce privilege from those that are truly transformative. There is no presumption here that reproducing privilege is unworthy – if that is what some colleges are in the business of doing they can do this more or less well, as in the case of Harvard (46 raw score, 44 adjusted score) vs. the University of Chicago (33 raw score, 28 adjusted score). Harvard’s mission is to train ‘citizen-leaders’ and it appears to be doing this well. It takes in enormously privileged students and produces economically successful graduates. The University of Chicago is an interesting comparison because it also takes in enormously privileged students but the outcomes of these students are much less what would be expected. It may be that the University of Chicago is much better at research than training its undergraduates for successful careers. If so, undergraduates should probably be aware of that.

Most public universities, on the other hand, have as their mission some notion of transforming the lives of the citizens of their respective states. Since UTEP ranked first in our sample, it may be worthwhile to take a look at their mission statement,

‘The University of Texas at El Paso commits itself to providing quality higher education to a diverse student population. Classified as a Doctoral/Research-Intensive university, UTEP seeks to extend the greatest possible educational

access to a region which has been geographically isolated with limited economic and educational opportunities for many of its people. The University **will ensure that its graduates** obtain the best education possible, one which is equal, and in some respects superior, to that of other institutions, so that UTEP's graduates **will be competitive in the global marketplace** [website, emphasis added]

UTEP includes both aspects of UTERM in its mission statement: access to a diverse set of students and economic returns. Based on its place in UTERM, it appears to be living up to its mission. Other public universities include similar statements in their missions but do not. Surely it makes more sense to compare these would-be-transformative institutions against each other rather than with those colleges whose mission is more broadly the reproduction of elites. But in the world of rankings, UTEP and all other such successful public universities who serve less privileged students ranks much lower than the reproducers of privilege.

In addition to the reproducers of privilege (RP) and truly transformative colleges (TT), we can categorize schools as safe bets (SB) or troubled returns (TR). Safe bets are those colleges which take in a fairly average set of students and graduate most of them to expected average outcomes. Most colleges can be seen to fall in this category. Troubled returns are those colleges whose students fare much less well than would be expected, regardless of LOP score. The cutoff for what ‘much less well’ is probably debatable but as a starting place there were four colleges in the sample whose adjusted scores were in the negative range. Negative scores could be red flags. It is impossible to score in the negative range without very high numbers of incompleters and loan defaults. If we widen our reach to C&U scoring in the lowest quintile, we see a range of institutional types in the troubled returns category, but the majority are liberal arts and private colleges with relatively large price tags. Table 7 highlights representative colleges in each of the four categories. Note that just as there are no elite colleges and universities in troubled returns, nor are they truly transformative.

Table 7. Examples of Truly Transformative Institutions, Reproducers of Privilege, Safe Bets and Troubled Returns

	ELITE/Flagships	Private LIBERAL ARTS	PUBLIC/HBCU Institutions
TRULY TRANSFORMATIVE (upper quintile)	UC Davis UCLA UT-Austin UC-Berkeley	Fairleigh Dickinson Bloomfield College La Salle U. Mercy College Colgate	UTEP CSU-LA CUNY – S.I Howard U. UC Davis
REPRODUCERS OF PRIVILEGE	Harvard Stanford MIT Brown Cornell	William and Mary C. of the Holy Cross Reed College Carleton College Kenyon College	-----

SAFE BETS	-----	Wofford College St. Thomas Aquinas Southwestern U. Hiram College Drew U.	UNC-Chapel Hill Oklahoma State Towson U. West Virginia U. IU-Bloomington
TROUBLED RETURNS (lowest decile)	-----	Guilford Rollins Drury U. Marymount Manhattan	UNC-Asheville Western Oregon U Boise State U.

One last point about these four categories must be mentioned. While students should probably avoid troubled returns, their choice between RP, SB, or TR institutions must remain a deeply personal one. A working-class student who does in fact gain admittance to Harvard may do well to accept rather than attending UTEP, despite UTEP’s transformative capabilities, simply because Harvard’s outcomes are, on average, much better than UTEP’s. Additionally, an individual’s aid package will alter the individual payoff calculus. UTEP is a transformative school because its outcomes are better than predicted by the type of students it admits.

The rankings are less helpful for individuals and more helpful as a national scorecard – pointing schools who rank low to reform ‘bad practices’ that are harming their students (wherever they may fall on the rankings system). It also alerts states to the great good that many of their publics are doing. Small changes – admitting more Pell Grant recipients, for example, or ensuring more timely completions of degrees can make substantial differences in the outcome scores. Thus, state legislatures can hold their public universities accountable. Not only does this allow comparisons between schools, but more importantly perhaps it can provide a measure of progress over time.

Finally, it is important to be clear that the Transformation Model is not designed to tell us what happens *inside* the university (although I would question what is actually being learned in schools that cost the same and serve the same populations and whose graduates default on loans and earn salaries below their peers). I think this is an advantage, however. Colleges are autonomous and teach according to different theories and practices. What works in a small college won’t work in a university serving 50,000 students - one reason why teacher to student ratios always undercount the impact of flagship universities. A one-size-fits-all ranking system, even when broken down between private and public, liberal arts colleges and national universities, will always obscure the special strengths of particular institutions. It is important to have a measure of economic returns, however, one that takes into account the different starting places of entering college students. This allows us not only to be critical of those schools that are failing their graduates, but also to reward the very high number of mostly public colleges and universities that are truly transforming the lives of their students.

Author Bio

Allison L. Hurst is an Associate Professor of Sociology in the School of Public Policy at Oregon State University. She is currently working on a project that explores the differences, acquisition, and deployment of social, cultural, and economic capital among college students attending two

comparable public universities as well as a sociological social history project that, re-examines the most economically egalitarian period in US history, 1948 to 1964. Allison's most recent publication (co-authored with Sandi Kawecka Nenga) is *Working in Class: How Our Social Backgrounds Affect Our Teaching, Scholarship, and Work in the Academy* (Rowman & Littlefield, 2016).

Bibliography

- Alon, S. 2007, 'Overlapping disadvantages and the racial/ethnic graduation gap among students attending selective institutions', *Social Science Research*, vol. 36, no. 4, pp. 1475-1499.
- Aries, E. and Seider, M. 2005, 'The interactive relationships between class identity and the college experience: the case of lower-income students', *Qualitative Sociology*, vol. 28, no. 4, pp.419-443.
- Arum, R. and Roksa, J. 2010, *Academically adrift: limited learning on college campuses*, University of Chicago Press, Chicago.
- Berg, G. A. 2010, *Low income students and the perpetuation of inequality: higher education in America*, Ashgate Pub, Burlington, VT.
- Black, D. A. and Smith, J. 2006, 'Estimating the returns to college quality with multiple proxies for quality', *Journal of Labor Economics*, vol. 24, pp. 701-728.
- Bowman, N. A. and Bastedo, M. 2009, 'Getting on the front page: organizational reputation, status signals, and the impact of *US News & World Report* on student decisions', *Research in Higher Education*, vol.50, pp. 415-436.
- Brand, J. E. and Halaby, C.N. 2006, 'Regression and matching estimates of the effects of elite college attendance on educational and career achievement', *Social Science Research*, vol. 35, no.3, pp. 749-770.
- Brand, J.E. and Xie, Y. 2010, 'Who benefits most from college?: evidence for negative selection in heterogeneous economic returns to higher education', *American Sociological Review*, vol. 75, no.2, pp. 273-302.
- Brooks, R. 2005, 'Measuring university quality', *The Review of Higher Education*, vol. 29, no.1, pp. 1-21.
- Crane, D. 1969, 'Social class origin and academic success: the influence of two stratification systems on academic careers', *Sociology of Education*, vol. 42, no.1, pp.1-17.
- Di Leo, J. R. 2011, 'Falling into rank: how I learned to stop worrying and love journal rankings', *Journal of Scholarly Publishing*, vol. 42, no.3, pp. 323-339.
- Ehrenberg, R. G. 2003, 'Reaching for the brass ring: *The U.S. News & World Report* rankings and competition', *The Review of Higher Education*, vol. 26, no.2, pp.145-162.
- Eff, A. 2012, 'Identifying the best buys in U.S. higher education', *Research in Higher Education*, vol. 53, no. 8, pp. 860-887.
- Engberg, M. E. 2012, 'Pervasive inequality in the stratification of four-year college destinations', *Equity and Excellence in Education*, vol. 45, no.4, pp. 575-59.
- Field, K. 2013, 'Obama plan to tie student aid to college ratings draws mixed reviews', *Chronicle of Higher Education*, August 22, viewed May 29 2019, <<https://www.chronicle.com/article/Obama-Plan-to-Tie-Student-Aid/141229>>
- Fitzgerald, R A. 2000, 'College quality and the earnings of recent college graduates', *NCES 2000-*

- 043, National Center for Education Statistics, US Department of Education, Washington, DC, viewed May 29 2019, < <https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2000043>>
- Garber, T. P. and Cheung, S.Y. 2008, 'Horizontal stratification in postsecondary education: forms, explanations, and implications,' *Annual Review of Sociology*, vol. 34, pp. 299-318.
- Gnolek, S., Falciano, V.T., and Kuncl, R.W. 2014, 'Modeling change and variation in *US News & World Report* college rankings: what would it really take to be in the Top 20?', *Research in Higher Education*, vol. 55, 7pp. 61-79.
- Goyette, K. A. and Mullen, A.L. 2006, 'Who studies the arts and sciences? social background and the choice and consequences of undergraduate field of study', *Journal of Higher Education*, vol. 77, no. 3, pp. 497-538.
- Grodsky, E. and Jackson, E. 2009, 'Social stratification in higher education', *Teachers College Record*, vol. 111, no.10, pp. 2347-2384.
- Horn, A. S. and Lee, G. 2016, 'The reliability and validity of using regression residuals to measure institutional effectiveness in promoting degree completion', *Research in Higher Education*, vol. 57, pp. 469-496.
- Hout, M. 2012, 'Social and economic returns to college education in the *US News & World Report*', *Annual Review of Sociology*, vol. 38, pp. 379-400.
- Hoxby, C. M. 2001, 'The return to attending a more selective college: 1960 to the present', in Devlin, M. and Myerson, J. (eds.), *Forum futures: exploring the future of higher education*, Jossey-Bass, San Francisco.
- Hu, S. & Wolniak, G. 2010, 'Initial evidence on the influence of college student engagement on early career earnings', *Research in Higher Education*, vol.8, pp. 750-766.
- Hussey, A.J., and Swinton, O.H. 2011, 'Estimating the ex ante expected returns to college', *American Economic Review*, vol. 101, no. 3, pp. 598-602.
- Kahlenberg, R.D. 2004, *America's untapped resource: low-income students in higher education*, Century Foundation Press New York.
- Kim, J., Kim J., Jaquette, O., and Bastedo, M.N. 2014, 'Institutional stratification and the post-college labor market: comparing job satisfaction and prestige across generations', *Journal of Higher Education*, vol. 85, no.6, pp. 761-789.
- Leparulo, P. 2015, 'Measuring the outcomes of Maine's college graduates', Report, *Federal Reserve Bank of Boston: Communities & Banking*, viewed May 29 2019, < <https://www.bostonfed.org/publications/communities-and-banking/2015/spring/measuring-the-outcomes-of-maines-college-graduates.aspx>>
- Liu, X., Thomas, S.L. and Zhang, L. 2010, 'College quality, earnings, and job satisfaction: evidence from recent college graduates', *Journal of Labor Research*, vol. 31, no. 2, pp. 183-201.
- Mattern, K. D., Shaw, E.J., and Kobrin, J. L. 2010, 'Academic fit: is the right school the best school or is the best school the right school?', *Journal of Advanced Economics*, vol. 21, no. 3, pp. 368-391.
- McDonough, P. M., Antonio, A.L., Walpole, M. and Perez, L.X. 1998, 'College rankings: democratized knowledge for whom?', *Research in Higher Education*, vol. 39, no. 5, pp. 513-537.
- Melguizo, T. 2010, 'Are students of color more likely to graduate from college if they attend more selective institutions?', *Education Evaluation and Policy Analysis*, vol. 32, pp. 230-248.
- Monks, J. 2000, 'The returns of individual and college characteristics: evidence from the NLSY', *Economics of Education Review*, vol. 19, pp. 279-289.

- Mullen, A. L. 2009, 'Elite destinations: pathways to attending an Ivy League university', *British Journal of Sociology of Education*, vol. 30, no. 1, pp. 15-27.
- Ostrove, J. M. and Long, S.M. 2007, 'Social class and belonging: implications for college adjustment', *The Review of Higher Education*, vol. 30, no. 4, pp. 363-389.
- Paulsen, M. B. 1998, 'Recent research on the economics of attending college: returns on investment and responsiveness to price', *Research in Higher Education*, vol. 39, no. 4, pp. 471-48.
- Pryor, J.H, Egan K., Blake, L.P., Hurtado, S., Berdan, J., and Case, M. H. 2012, *The American freshman: national norms fall 2012*, Higher Education Research Institute, UCLA, Los Angeles.
- Reynolds, O. E. 1927, *The social and economic status of college students*, Teachers college, Columbia university, New York.
- Roksa, J. and Levey, T. 2010, 'What can you do with that degree? college major and occupational status of college graduates over time', *Social Forces*, vol. 89, no. 2, pp. 389-415.
- Sibulkin, A. 2011, 'Diverse colleges of origin of African American doctoral recipients, 2001-2005: historically Black colleges and universities and beyond', *Research in Higher Education*, vol. 52, no. 8, pp. 830-852.
- Stuber, J. M. 2011, *Inside the college gates: how class and culture matter in higher education*, Lexington Books, Lanham, Md.
- Thomas, S. L. and Zhang, L. 2005, 'Post-baccalaureate wage growth within four years of graduation: the effects of college quality and college major', *Research in Higher Education*, vol. 46, no. 4, pp. 437-459.
- Torche, F. 2011, 'Is a college degree still the great equalizer? intergenerational mobility across levels of schooling in the US', *American Journal of Sociology*, vol. 117, no. 3, pp. 763-807.
- Volkwein, J.F. and Grunig, S.D. 2005, 'Resources and reputation in higher education: double, double, toil and trouble', in J.C. Burke et al. (ed.), *Achieving accountability in higher education*, Jossey-Bass, San Francisco, pp.246-274.
- Walpole, M. 2007, 'Economically and educationally challenged students in higher education: access to outcomes', *ASHE Higher Education Report*, vol. 33, no. 3, pp. 1-113.
- Winston, G. C. and Hill, C.B. 2005, 'Access to the most selective private colleges by high-ability, low-income students: are they out there?', *Williams Project on the Economics of Higher Education DP-69*, Department of Economics, Williams College, viewed May 29 2019, <<https://ideas.repec.org/s/wil/wilehe.html>>
- Zhang, L. 2008, 'Gender and racial gaps in earnings among recent college graduates', *Review of Higher Education*, vol. 32, no. 1, pp. 51-72.