College Rankings

History, Criticism and Reform

Luke Myers and Jonathan Robe

Center for College Affordability and Productivity



A Report from the Center for College Affordability and Productivity

March 2009

About the Authors

Luke Myers is a senior studying political science through the Honors Tutorial College at Ohio University. He is currently writing an undergraduate thesis on deliberative democracy and has been a research assistant with CCAP since June 2008.

Jonathan Robe has been a research assistant at CCAP since August 2007. He is an undergraduate student at Ohio University majoring in mechanical engineering and currently serves as an Engineering Ambassador for the college.

About the Center for College Affordability and Productivity

The Center for College Affordability and Productivity is a nonprofit research center based in Washington, DC, that is dedicated to research on the issues of rising costs and stagnant efficiency in higher education.

1150 17th ST. NW #910 202-375-7831 (Phone) Washington, DC 20036 202-375-7821 (Fax)

www.centerforcollegeaffordability.org www.collegeaffordability.blogspot.com

Table of Contents

Introduction	5
The History of Academic Quality Rankings	5
Contributions and Criticisms of College Rankings	22
Effects of College Rankings	28
College Rankings Reform	31
Conclusion	38
Figures and Tables	
Table 1: Correlations Between American MBA Rankings	24
Figure 1: Correlation Between USNWR Ranks with	
Previous Year's Ranks (National Universities)	31
Figure 2:Correlation Between USNWR Ranks with	
Previous Year's Ranks (Liberal Arts Colleges)	31
Table 2: Correlations of Component Ranks to Overall Rank in	
U.S. News (National Universities)	32
Table 3: Correlations of Component Ranks to Overall Rank in	
U.S. News (Liberal Arts Colleges)	32
Table 4: Dependent Variable is the Ranking Score,	
Ordinary Least Squares Estimation	38
Table 5: Dependent Variable is the Ranking Score,	
Ordinary Least Squares Estimation	39
Table 6: Dependent Variable is the Ranking Score,	
Ordinary Least Squares Estimation	40

Introduction

Today, college quality rankings in news magazines and guidebooks are a big business with tangible impacts on the operation of higher education institutions. The college rankings published annually by *U.S. News and World Report (U.S. News*) are so influential that Don Hossler of Indiana University derisively claims that higher education is the victim of "management" by the magazine. There is certainly support for such a claim: college rankings—particularly those of *U.S. News*—sell millions of copies when published, affect the admissions outcomes and pricing of colleges, and influence the matriculation decisions of high school students throughout the world.¹

How did academic quality rankings of colleges and universities become so powerful in higher education? A review of their historical development in the first section of this study may surprise many readers. While college professors and administrators alike largely decry rankings today, their origin lies in academia itself. Begun as esoteric studies by lone professors, college rankings' development into the most popularly accepted assessment of academic quality was fueled by the very institutions of higher education they now judge. While the purpose and design of academic quality rankings has evolved during the century since their creation, their history teaches one clear lesson: college rankings fill a strong consumer demand for information about institutional quality, and as such, are here to stay for the foreseeable future.

Various approaches to college rankings have different benefits and each is subject to legitimate criticism, all of which should be seriously considered in light of the powerful effects that a widely-distributed ranking can have on institutions of higher education and the students seeking to enter them. Sections II and III will explore these aspects of college rankings, respectively. In light of the historical lessons revealed in Section I, however, movements that seek to reform college rankings should be focused on producing *better* rankings, rather than on trying to eliminate or ignore them. Section IV will survey multiple new indicators of academic quality that many view as potential improvements over the indicators upon which current college rankings are based.

The History of Academic Quality Ranking

Many and various efforts have been made to assess the quality of higher education institutions. Accreditation agencies, guidebooks, stratification systems, and rankings all have something to say about the quality of a college or university but express it in very different ways. For clarity, we will adopt higher education researcher

David Webster's definition of "academic quality rankings." For Webster, an academic quality ranking system has two components:

- 1. It must be arranged according to some criterion or set of criteria which the compiler(s) of the list believed measured or reflected academic quality.
- 2. It must be a list of the best colleges, universities, or departments in a field of study, in numerical order according to their supposed quality, with each school or department having its own individual rank, not just lumped together with other schools into a handful of quality classes, groups, or levels.²

All but one of the studies and publications discussed below will fit both criteria and so will qualify as "academic quality rankings."

Ranking systems that meet these two criteria can be further distinguished by their placement within three polarities. First, some rankings compare individual departments, such as sociology or business, within a college or university, while others measure the quality of the institutions as a whole, without making special note of strong or weak areas of concentration. Second, rankings differ by whether they rank the quality of graduate or undergraduate education. The judging of graduate programs and comparing of individual departments are often coupled together in a ranking system. This should come as little surprise considering the specialization of graduate-level education. Similarly, ranking undergraduate education usually, but not always, involves ranking whole institutions, probably due to the fact that a well-rounded education is often viewed as desirable at this level.

More important than *what* rankings judge is *how* they do the judging. Most academic quality rankings to this point have used one of two primary strategies for determining quality: outcomes-based assessment or reputational surveys, although other objective input and output data such as financial resources, incoming student test scores, graduation rates, and so forth have often been used to supplement these primary measures. Rankings that look at college outcomes are often concerned with approximating the "value-added" of a college or university. They use data about students' post-graduate success, however defined, to determine the quality of higher education institutions and have often relied on reference works about eminent persons such as *Who's Who in America*. Reputational rankings are those which are significantly based on surveys distributed to raters who are asked to list the top departments or institutions in their field or peer group.

Either form of academic quality rankings—outcomes-based or reputational—can be used in departmental or institutional rankings and graduate or undergraduate rankings. In fact, there have been two major periods in which each method of ranking was ascendant: outcomes-based rankings, derived from studies of eminent graduates, were published in great number from 1910 to the 1950s, while reputational rankings became the norm starting in 1958 and continuing to the present. While there has been some renewed interest in outcomes-based rankings recently, they have yet to regain parity with reputational rankings in terms of popularity. The rest of this section will examine a number of major academic quality rankings throughout history, and explore their development from esoteric studies into one of the most powerful forces in higher education.

Early Outcomes-Based Rankings

The first college rankings developed in the United States out of a European preoccupation—especially in England, France, and Germany—with the origins of eminent members of society. European psychologists studied where eminent people had been born, raised up, and attended school in an attempt to solve the question of whether great men were the product of their environment (especially their university) or were simply predestined to greatness by their own heredity. In 1900, Alick Maclean, an Englishman, published the first academic origins study entitled *Where We Get Our Best Men*. Although he studied other characteristics of the men, such as nationality, birthplace, and family, at the end of the book he published a list of universities ranked in order by the absolute number of eminent men who had attended them. In 1904, another Englishman, Havelock Ellis—a hereditarian in the ongoing nature versus nurture debate—compiled a list of universities in the order of how many "geniuses" had attended them.⁴

In each case, neither author explicitly suggested the use of such rankings as a tool for the measurement of the universities' quality. Although there seems to be an implicit quality judgment in simply ranking universities according to their number of eminent alumni, the European authors never made the determination of academic quality an explicit goal. However, when Americans began producing their rankings with this very aim, they used similar methodologies and data. Many of the earliest academic quality rankings in the United States used undergraduate origins, doctoral origins, and current affiliation of eminent American men in order to judge the strengths of universities.⁵

The first of these rankings was published by James McKeen Cattell, a distinguished psychologist who had long had an interest in the study of eminent men. In 1906, he published *American Men of Science: A Biographical Dictionary*, a compilation of short

biographies of four thousand men that Cattell considered to be accomplished scientists, including where they had earned their degrees, what honors they had earned, and where they had been employed. He "starred" the thousand most distinguished men with an asterisk next to their biography. In the 1910 edition of *American Men of Science*, Cattell updated the "starred" scientists and aggregated the data about which institutions these men had attended and where they taught at the time, giving greater weight to the most eminent than to the least. He then listed the data in a table with the colleges in order of the ratio of this weighted score to their total number of faculty, thereby creating the first published academic quality ranking of American universities.⁶

Cattell understood that he was making a judgment about these institutions' quality as evidenced by his titling the table "Scientific Strength of the Leading Institutions," and his claim that "[t]hese figures represent with tolerable accuracy the strength of each institution." Cattell was also aware that prospective students would be interested in the judgments of quality. He wrote, "Students should certainly use every effort to attend institutions having large proportions of men of distinction among their instructors." Furthermore, Cattell claimed that the "figures on the table appear to be significant and important, and it would be well if they could be brought to the attention of those responsible for the conduct of the institutions," implying a belief that the rankings represented a judgment of quality that could be improved over time if the institutions took the correct actions.⁷

Although Cattell's first study was not based purely on the measured outcomes of the institutions he ranked, it was central to the development of later outcomes-based rankings. Cattell himself would continue to publish similar studies in which he judged institutions of higher education based on the number of different eminent people—not just scientists—they both produced and employed, without ever fundamentally altering his methodology. From Cattell's 1910 study until the early 1960s, the quality of institutions of higher education would be most frequently judged using this method of tracking the educational background of distinguished persons.⁸

One researcher who was greatly influenced by Cattell's work, but who even more explicitly dealt with the quality of academic programs, was a geographer from Indiana University named Stephen Visher. Interested in why geographical areas demonstrated a disparity in the number of scientific "notables" they produced, Visher looked at the undergraduate education of the 327 youngest "starred" scientists in Cattell's 1921 edition of *American Men of Science*. Such an approach tested the hypothesis that the disparities resulted because "the leaders come from those who are greatly stimulated in colleges." He ranked the top seventeen institutions by the ratio of the young "starred" scientists to total student enrollment, thereby creating the

first enrollment-adjusted outcomes-based ranking. Visher suggested that the rank demonstrated the "comparative success of these institutions in inspiring undergraduate students," and argued that "[t]he conspicuous contrasts...in the number and percentage of graduates who later become leaders suggest that there are marked variations in the stimulating value of institutions."9

Beverly Waugh Kunkel, a biologist at Lafayette College, and his co-author Donald B. Prentice, then president of the Rose-Hulman Institute of Technology, repeatedly used a methodology similar to that of Cattell and Visher, but stated their interest in the academic quality of universities even more explicitly. In their first study, published in 1930, Prentice and Kunkel expressed interest in "what elements constitute a successful institution," especially in light of the large investments that individuals were making in their educations. The authors believed that "undoubtedly the most reliable measure" of a higher education institution was "the quality of product." Therefore, Prentice and Kunkel measured academic quality by the number of a college's undergraduate alumni listed in *Who's Who in America.*¹⁰

Kunkel and Prentice repeated essentially the same methodology in periodical studies from 1930 to 1951. They ranked schools according to the number of baccalaureate-earning alumni who were listed in *Who's Who.*¹¹ In the 1930 study, the authors provided a table ranking the schools by absolute number of graduates listed and a second table ranking them according to the percentage of a school's living alumni who were listed. The authors noted that an overrepresentation of ministers and college professors and an underrepresentation of engineers in *Who's Who* likely skewed the results of their rankings. In the 1951 study, the authors listed the schools alphabetically with the absolute number of alumni listings and their numerical rank. This later study did not include a percentage based ranking, but instead focused on the time period from which the listed alumni graduated, hoping that this might be of use in identifying good practices for those familiar with an institution's historical operations.¹²

One final early study that deserves mention is the first and last attempt by the federal government to explicitly compare academic quality among institutions. In 1910, the American Association of Universities (AAU) asked Kendric Charles Babcock, Higher Education specialist in the Bureau of Education, to publish a study of the undergraduate training at colleges so that graduate schools would be able to know which applicants were best prepared. The Bureau of Education was chosen because it was believed by AAU that the rankings would be more widely accepted if they were compiled by an impartial source without a connection to a university.

Babcock's study was a stratification and not a ranking. When finished, he divided 344 institutions into four different classes rather than supplying an individual rank to each school. As with most of the early studies mentioned above, Babcock measured quality based on the outcomes an institution produced—here, the performance of schools' graduates after they entered graduate school—but he was not greatly influence by Cattell's quantitative, eminent person methodology. On visits to several graduate schools, Babcock "conferred with deans, presidents, and committees on graduate study," and "inspected the credentials and records of several thousands of graduate students...in order to ascertain how such students stood the test of transplanting." 14

The accidental release of a draft of the study to the newspapers resulted in such a furor from the deans and presidents of colleges classified lower in the rankings that President Taft issued an executive order prohibiting the study's official release. The Commissioner of Education, P.P. Claxton, tried to soothe the disgruntled by admitting that the classification was "imperfect" because its single criterion of graduate school performance failed to account for the fact that many colleges may perform very highly in their provision of services to those students who do not go on to graduate school. Neither Claxton's explanations nor the praise the classification received from some deans and presidents (mostly from class I schools), were enough to convince President Wilson from rescinding Taft's order when asked to do so by AAU upon his arrival in the White House. This historic episode demonstrates one reason why the federal government has never attempted to rank or in any way judge the comparative academic quality of higher education institutions since.

The Rise of Reputational Rankings

Reputational surveys would become the predominant method for producing academic quality rankings beginning in 1959, with the most popular ranking today, one by *U.S. News and World Report*, containing a strong component of reputational evaluation. However, this methodology was developed much earlier, in 1924, by Raymond Hughes, a chemistry professor at Miami University in Ohio. When asked by the North Central Association of Schools and Colleges to complete a study about graduate school quality, Hughes would turn to the opinion of his fellow faculty instead of relying on the then popular outcome-based methodology. ¹⁶

Hughes circulated two requests to Miami University faculty in twenty fields of study—nineteen liberal arts disciplines and the professional discipline of education. The first sought from each faculty member a list of forty to sixty instructors who taught their discipline in American colleges and universities. The second request asked the recipients to rate, on a scale of one to five, the departments of thirty-six

institutions that offered a degree in their discipline, so as to create "a list of the universities which conceivably might be doing high grade work leading to a doctor's degree." ¹⁷

Hughes received about a 50 percent response rate. After weighting the ratings, Hughes produced a table that listed the departments according to how many 1, 2, 3, 4 or 5 ratings they had received. Although he did not calculate an overall score for each department, the ordered list, based on a specific criterion, meets the definition of an academic quality rating, the first such list determined by a department's reputation among selected raters. Hughes also did not aggregate the ranks of the departments to form an institution-wide ranking.

During his chairmanship of the American Council on Education, Hughes would publish another study in 1934 on graduate school quality of much wider scope. Hughes's second study was an improvement over the first in many respects. First, the 1934 study covered thirty-five disciplines as opposed to the twenty disciplines in his earlier study. The second study also gathered opinions from a more diverse field of respondents; to compile his list of raters, Hughes asked the secretary of each discipline's national society for a list of one hundred scholars that would fully represent the field and its sub-fields, resulting in a greater number of respondents for each discipline. However, while the 1934 study helped to refine the reputational methodology, it was not a ranking. Instead of listing the departments in order of their rating, Hughes simply listed alphabetically any department that at least half of the raters had judged as adequate.¹⁸

Though developed during the same period as other outcomes-based rankings, the reputational methodology of judging academic quality was largely absent for twenty-five years after Hughes's 1934 study. It would reappear in the appendix of *Graduate Study and Research in the Arts and Sciences at the University of Pennsylvania*, published by a humanities professor, Hayward Keniston. The ranking was compiled in connection with work he was doing for the University of Pennsylvania in 1959 to help compare it to other American research universities. Although Keniston's ranking did not gather much attention beyond the walls of his institution, its publication nonetheless marks the beginning of the decline of outcomes-based rankings and the rise of reputation-based rankings, a shift that would be complete a decade later.

The ranking in Keniston's study relied solely on the opinions of twenty-four department chairpersons at each of twenty-five top universities. The universities from which the raters came were chosen based on their membership in the Association of American Universities, the number of doctorates granted, and their geographical distribution. Keniston was only interested in comprehensive research universities com-

parable to the University of Pennsylvania, so schools such as Massachusetts Institute of Technology and California Institute of Technology were not included due to their technical nature, and Michigan State and Penn State were not included because of their limited Ph.D. programs.¹⁹

Once Keniston identified the raters, he asked them to rank the fifteen strongest departments in their discipline at the twenty-five universities to which he had sent surveys. After an 80 percent response rate, resulting in about twenty different rankings per discipline, Keniston aggregated the rankings of departments into the four broad categories of humanities, social sciences, biological sciences and physical sciences. He also aggregated the disciplinary ratings into institution-wide rankings, making his study the first institution-wide ranking determined through reputational surveys. It should be noted that Keniston's choice of which disciplines to include seems to have been influenced by a desire to improve the University of Pennsylvania's position in the rankings. Eleven of the twenty-four came from the humanities, including Oriental studies and Slavic studies, in which the University of Pennsylvania ranked eighth and sixth respectively—the university's two highest ranks overall—and did not include engineering, one of the university's less prestigious departments.²⁰

From 1959 to 1966, the reputational methodology quietly gained ground in the world of academic quality rankings. After Keniston, five reputational rankings were completed, one unpublished, but none of which received any special attention. An Australian geographer published a ranking of American geography departments in an Australian journal in 1961. Albert Somit and Joseph Tanehaus published a book -length study of American political science departments in which they ranked the top thirty-three graduate departments in 1964. In 1966, Sam Sieber (with the collaboration of Paul Lararsfeld) ranked departments of education according to raters' views on their research value, subscribers to *Journal of Broadcasting* ranked broadcasting graduate programs, and Clark Kerr, then president of the University of California, created an unpublished ranking of medical schools affiliated with universities who belonged to AAU. During this time, however, outcomes-based rankings had not yet disappeared; the well-known psychologist Robert Knapp was still publishing rankings based on academic origins up until 1964.²¹

The ascendancy of reputational rankings can be said to have truly started with the methodological advances of Allan Cartter, who published the 1966 Assessment of Quality in Graduate Education (Cartter Report). The Cartter Report ranked twentynine disciplines, similar to Hughes and Keniston's studies, but it was an improvement over these reputational rankings in several significant ways. First, it polled senior scholars and junior scholars in addition to department chairpersons, result-

ing in almost 140 rankings per discipline, providing a more diverse and larger body of opinions than both previous rankings. Second, Cartter had raters rank the disciplines at 106 different institutions, more than any previous reputational ranking. Finally, the *Cartter Report* ranked the departments according to the two criteria of "quality of the graduate faculty" and "rating of the doctoral training program," instead of just one criterion, as in both Hughes studies and in Keniston's ranking.

The respondents rated each department on a scale of one to five. In addition to ranking the departments, Cartter stressed his interest in quality by providing labels based on their scores. All those ranked 4.01 and up were labeled "distinguished" and those ranked 3.01 to 4.00 were labeled "strong." For departments scoring from 2.01 to 3.01, Cartter listed them alphabetically, split in the middle between those labeled "good" and "adequate plus." Although Cartter did not aggregate his departmental ratings into institution-wide rankings, three other authors performed the task with his data after its publication. Cartter also provided more analysis of his own rankings than any previous reputational study, including geographical distribution of highest ranked departments, relationship between ranking and faculty compensation, and the correlation between faculty publications and their score for "quality of graduate faculty."²³

Cartter's ranking not only had the most comprehensive methodology to date, it also enjoyed the best critical reception. It received mass attention by earning more reviews than any previous reputational ranking study, most of which were positive. In addition to praise from higher education officials, magazines such as *Time* and *Science* lauded the assessment provided by the study. Once published, the report sold approximately 26,000 copies.²⁴ This commercial success and critical acclaim can be understood as one of the prime reasons reputational rankings became the overwhelming norm after 1966.

In 1970, Kenneth Roose and Charles Andersen sought to replicate Cartter's study, although with a self-admitted goal of de-emphasizing the "pecking-order" of the first ranking. Roose and Andersen's *A Rating of Graduate Programs* rounded departments' ratings to one decimal rather than two, resulting in more ties. Yet only the rank, and not these scores, were published. There were no descriptive labels assigned to a program's score and the Roose/Andersen study only provided the ordinal rank of the departments based on their "faculty quality" score. For the "program effectiveness" score, Roose and Andersen simply listed departments in order of their scores without including an ordinal position.²⁵

Additionally, the Roose/Andersen study expanded the number of disciplines included to thirty-six, the number of institutions at which these disciplines were

ranked to 130, and the number of usable responses to approximately 6,100 (from 4,000 in the *Cartter Report*). Like Cartter's study, Roose and Andersen did not aggregate the departmental ratings into institution-wide rankings, but publications such as *Newsweek* and the *Journal of Higher Education* did. Despite this expanded scope, decreased importance of the "pecking-order," and significant press coverage, the Roose/Andersen study did not receive the same reception as did the *Cartter Report* four years earlier. Indeed, it received much criticism from academics, one of whom complained that the rankings did not reflect recent increases in quality at newer and developing universities.²⁶

When published in 1982, the Assessment of Research-Doctorate Programs in the United States (Assessment), produced by the National Academy of Sciences in conjunction with the National Research Council, was the largest academic quality ranking project ever undertaken. The Assessment rated a total of 2,699 programs at 228 different institutions and provided detailed data about hundreds of these programs. The entire study cost more than \$1.1 million in 2008 dollars before it was completed. Unfortunately, the large amounts of information collected were not presented in an easily understandable fashion. For each measure of quality, the programs were listed alphabetically by institution with their standardized scores listed next to them. Since several dozen programs were listed for most disciplines, each rank was difficult to determine. Furthermore, no attempt was made by the authors of the study to aggregate these scores into institution-wide rankings.²⁷

In addition to being the largest, the *Assessment* was also the first major reputational study to include non-reputational measures as well. Of the sixteen measures used, four were reputational, while the others covered areas such as program size, characteristics of graduates, library size, research support and publication records. However, the reputational measures were the most widely reported in news outlets, from the *Chronicle of Higher Education* to the *Washington Post*. Of twelve scholarly works about the NAS *Assessment*, all mentioned one or more reputational measure while only three discussed any of the other twelve measures. All of the twenty-nine news articles covering the study reported the ranks of programs according to at least one of the reputational measures but only seven reported the rankings according to any non-reputational measure.²⁸

A follow-up study produced by the National Research Council in 1995 studied much of the same data as the *Assessment*, but with an increased scope. The *Research-Doctorate Programs in the United States: Continuity and Change* ranked forty-one disciplines at 274 institutions, resulting in a total of 3,634 programs, costing \$1.7 million in 2008 dollars. The 1995 report improved on the reporting method of the earlier *Assessment* by publishing the names of institutions in rank order for each disci-

pline, although it did not provide the numerical rank in the tables. It used twenty measures of quality: three reputational, eight related to faculty research, two related to doctoral students and seven related to Ph.D. recipients. Once again, although many different quantitative measures were reported in the study, the rankings based on the reputational measurements received the most attention.²⁹ The next version of this graduate program ranking is expected to be officially released by the National Research Council shortly.³⁰

Re-Emergence of Undergraduate Rankings

The rise of reputational rankings starting in 1959 coincided with a decreased interest in ranking both institutions and programs for undergraduate study. Although undergraduate study was the primary concern of many of the early rankings, such as the unpublished Bureau of Education stratification and the outcomes-based, academic origins studies of Visher and Kunkel and Prentice, all of the reputational studies mentioned above focused on graduate-level programs. However, just as the first reputational studies were published during the ascendancy of outcomes-based rankings, a few reputational undergraduate rankings were published during the dominance of reputational graduate studies from the late 1950s to the early 1980s.

The first undergraduate ranking compiled through a reputational methodology was a little known study published in 1957 by journalist Chesly Manly in the *Chicago Sunday Tribune*. In an article, Manly ranked the top ten universities, coeducational colleges, men's colleges, and women's colleges according to their undergraduate quality, in addition to the top law and engineering graduate schools. The ranking was also the first with a reputational component to rate whole institutions rather than just departments or disciplines. Although the methodology was never fully disclosed, Manly claimed to have based the rankings on the opinions of a few dozen "consultants" in addition to a "great mass" of quantitative data.³¹

Another ranking that focused on both undergraduate and graduate rankings is *The Gourman Report*, authored by Jack Gourman. These reports began evaluating colleges in 1955 and Gourman began publishing the results in 1967, the latest of which appeared in 1997. Although the ranking offers a rating to two decimal places for each school included, the methodology used is almost completely unknown. Gourman has said that the final score is calculated by averaging scores on ten factors, which have more to do with administration policies than with any typical measure of reputation or student success. He has never revealed exactly what data were gathered, who and how they gathered it, or how it was compiled into the precise final score. Additionally, experts have claimed that the many one-hundredths of a point differences between schools with no large gaps is almost statistically impossi-

ble. Regardless of the opacity and questionable results, however, the Gourman ratings have historically been used in many scholarly reports, especially by economists interested in the relationships between college quality and other variables such as alumni earnings and student choice.³²

The first post-1959 undergraduate ranking developed with methodological rigor was a pilot study by Lewis Solmon and Alexander Astin, published in 1981. They provided raters from four states—California, Illinois, North Carolina, and New York—a list of 80 to 150 departments in each of seven different fields: biology, business, chemistry, economics, English, history, and sociology. They asked the raters to choose the best quality department from this list in their field and, in addition, to judge each department on six criteria: overall quality of undergraduate education, preparation for graduate school, preparation for employment, faculty commitment to teaching, scholarly accomplishments of faculty, and innovativeness of curriculum.³³

Solmon and Astin found that a rater's overall opinion of an undergraduate department was highly correlated with his or her opinion about the department's "scholarly excellence of faculty" and "commitment to undergraduate teaching." The first criterion was also highly correlated to whether the institution had a strongly-rated graduate program, which resulted in the top undergraduate programs appearing largely as a reflection of Cartter and Roose/Andersen's studies of top graduate programs. Since Solmon and Astin were interested in providing a new list of undergraduate quality, rather than simply repeating the same schools as those earlier studies, they excluded any institution that had received a high rating in the Roose/Andersen study from the results they reported in *Change* magazine. With those institutions left, they ranked departments in each field according to how many times each ranked in the top ten of the six criteria. In a second table, they listed institutions by how many of their departments had been listed in the top ten in their respective disciplines. One interesting result was that some typically highly-regarded colleges were ranked among the top in fewer than five of the disciplines.³⁴

Rankings in Popular Publications

The face of academic quality rankings would be revolutionized with the undergraduate reputational ranking first published in 1983 by *U.S. News and World Report*. Even though authors of academic quality rankings from as early as Cattell—the very first—have noted the interest that prospective students might have in knowing the ranks of different institutions and departments, rankings would not play a large role in helping high school students choose what college to attend until *U.S. News* began publishing "America's Best Colleges." Until then, academic quality rankings were the province of professors and higher education administrators. Published as studies by

researchers in little-circulated academic books and journals, few college-bound students would have been able to find them, and the rankings were often too obscure to be very helpful even if located. Compiled by editors and published in a highly-circulated news magazine, the *U.S. News* rankings became the most widely read and more influential than any ranking that had come before them.³⁵

The first three *U.S. News* rankings were entirely based on reputational surveys. The schools were broken into categories according to their Carnegie classification, and college presidents were asked to name the top undergraduate institutions. In 1983, 1308 presidents were sent surveys and *U.S. News* received about a 50 percent response rate. The ranking was repeated two years later in 1985. The raters were once again college presidents, but in this ranking they were asked to select the top five undergraduate schools from a provided list of institutions similar to their own. For the third ranking in 1987, the surveyed presidents would be asked to rate the top ten institutions. By the third ranking, the response rate had increased to 60 percent, but the criticism from academia had gained momentum as well. A number of the presidents who had refused to respond argued that neither they nor their fellow presidents had the ability to judge the academic quality of any institution except their own.³⁶

The rankings in 1988 departed from the first three rankings in response to the increased criticism. Reportedly in consultation with college presidents and academic experts, *U.S. News* made two major methodological changes. First, they surveyed the opinions of academic deans and admissions officers in addition to those of college presidents, arguing that this would more adequately cover differing conceptions of quality. Second, they reduced the reputational component to just 25 percent of the overall ranking and determined the remaining 75 percent of a school's score using objective input and output data such as admissions selectivity, faculty strength, educational resources, and graduation rates.³⁷ Additionally, 1988 marked the beginning of the annual publishing of the ranking, as well as the year that *U.S. News* began publishing a book length college guide, *American's Best Colleges*, which included further in-depth information about the schools included in the rankings.³⁸

Over the past two decades, the methodology for the *U.S. News* undergraduate rankings has been through numerous incremental adjustments. In 1995, respondents were asked to consider the quality of undergraduate teaching in their ratings and the weight placed on the graduation rate was increased, due to a greater concern for measuring outcomes.³⁹ This interest in outcomes-based measurements further developed into what *U.S. News* called a "value-added" criterion in 1996, in which the ranking used a model to predict a school's expected graduation rate, using input factors such as test scores of incoming students, and then compared the predicted

to its actual graduation rate. The higher the actual rate was above the predicted rate, the higher the school's rank.⁴⁰

In 1999, *U.S. News* began to standardize the data used in the calculation of the rankings in order to bring their process more in line with accepted statistical practices. This allowed the outcomes to reflect the size of differences between schools in each component of the rankings, rather than simply the respective ranks. In 2000, they further tweaked their calculations by adjusting for the ratio of graduate students to undergraduate students at each school in order to eliminate a bias towards schools that spend large amounts of research money that primarily benefit graduate students rather than undergraduates. Related to methodological changes, the reliance on the Carnegie classifications to break schools into ranking categories has resulted in seemingly drastic changes for some schools from one year to the next. For example, Carnegie has updated their classifications twice since 2000. This means that schools can suddenly appear or vanish in one ranking category, even though there has been no significant change in their characteristics.

Some observers have attacked the magazine for these constant changes in methodology, arguing that they produce changes in some schools' ranks that are only reflective of the adjustments in the ranking's weighting rather than any true change in quality at the institution. However, college rankings scholar David Webster has applauded *U.S. News* for its constant tinkering as a demonstration of their receptiveness to criticism and constant striving for improvement. Indeed, in 1992, Webster considered *U.S. News*'s undergraduate rankings "by far the best such rankings ever published." An additional criticism, however, is that there have been many reports over the years of schools deliberately "fudging" their data or taking non-quality related steps in order to increase their *U.S. News* rank. 42

Two smaller but still relatively widely-circulated rankings deserve mention at this point. Shortly after *U.S. News* began publishing their undergraduate ranking annually, *Money* released its first annual "America's Best College Buys" in 1990. Rather than judging schools only by their *quality*, however, *Money* sought to rank schools according to their *value*—that is, the amount of quality per dollar of tuition spent. *Money* used statistical analysis to determine how much a college should be expected to cost based on a number of factors: test scores and class ranks of incoming students, faculty resources and quality, library resources, graduation and retention rates, and academic and career success of alumni. The schools were then ranked according to how far their actual "sticker price" tuition is below their predicted tuition. Beyond this, little more is known about *Money*'s methodology. One interesting result of ranking by value, however, was that *Money*'s top schools were highly diverse in their academic quality, price, and public/private affiliation.⁴³

In *The Best 368 Colleges*, The Princeton Review publishes numerous top-twenty rankings of various categories based on surveys completed by students at the 368 institutions profiled in the book. The edition published in 2008 contained sixty-one different rankings in the eight overall categories of academics, politics, demographics, quality of life, parties, extracurricular activities, schools by type, and social life. The rankings are determined by an eighty-one-question survey that students fill out about their own campuses. The respondents are nearly or completely self-selected. According to the Princeton Review website, in recent years 95 percent of the surveys have been filled out electronically by those students who sign up online, and they can be completed "anytime, anywhere." While clearly the least methodologically rigorous of their popular counterparts, the Princeton Review's rankings nonetheless earn significant media attention, and with the name of a major college-prep company behind them, almost certainly influence high school students' enrollment decisions.⁴⁴

The most recent academic quality ranking published by a popular news magazine was in a few respects a return to origins. Although the 2008 study was released on the website of Forbes magazine, it was largely developed by an academic—Richard Vedder, an economist at Ohio University and the director of the Center for College Affordability and Productivity.⁴⁵ In contrast to the input-heavy *U.S. News* ranking, Vedder's ranking was largely based on academic outcomes. Measurements such as the enrollment-adjusted number of alumni listed in Who's Who in America and the number of faculty winning national and international awards reflect the method of some of the very earliest academic quality rankings, such as Cattell, and Kunkel and Prentice. Additionally, the ranking measured student opinions of their professors as self-reported on Ratemyprofessors.com, the amount of debt students held at graduation, the colleges' four-year graduation rate, and the number of students winning nationally competitive awards. According to the article accompanying the rankings, Vedder's aim for the *Forbes*/CCAP ranking was to judge colleges from a student's perspective. The ranking's methodology was designed to account for students' primary concerns, such as the quality of undergraduate teaching and the outcomes of an education in terms of graduation, debt, and post-graduation career success.⁴⁶

Popular news magazines have not limited themselves to ranking undergraduate education. Only a few years after publishing their first undergraduate ranking, *U.S. News and World Report* produced their first ranking of top professional schools in 1987. This first edition ranked medical, law, engineering, and business graduate programs based completely on the opinions collected from surveys sent to deans of departments in those same fields.⁴⁷ Similar to the way their undergraduate rankings developed, the graduate and professional school rankings by *U.S. News* have ex-

panded and incorporated more objective data over the years. The most recent edition ranks institutions in twenty-two different professional and graduate fields. The rankings of the five professional fields (the original four fields plus education) each have a unique methodology, but they all include some measure of reputation among peers and employers and a measure of selectivity (in the form of schools' acceptance rates and entering students' test scores, GPA, etc.). Other factors used include faculty resources, job placement rates, and research strengths, though none of these factors is used in all five fields. The rankings of the other seventeen graduate disciplines—which can be grouped into six umbrella fields of sciences, library studies, social sciences and humanities, health, public affairs, and fine arts—are still completely determined by reputational surveys of department heads and directors of graduate study in the relevant discipline.⁴⁸

U.S. News appears to enjoy the same dominance in the graduate school ranking business that they have in the undergraduate rankings. Although some individuals and professional organizations produce alternative rankings—such as law professor Brian Leiter's Law School Rankings—no other regular large-circulation publication provides rankings covering the number of graduate and professional programs included by *U.S. News.*⁴⁹ The graduate school rankings produced by the National Research Council are by far the most comprehensive of such rankings and perhaps receive the most attention in the academic community, but their irregular publication (1982, 1995, and forthcoming) and convoluted organization of information (see above discussion) have limited their popular influence.

The one exception to *U.S. News*'s supremacy is in the field of business. Numerous major publications dealing with finance or business produce their own rankings of business schools or MBA programs. *Business Week, Forbes*, and the *Wall Street Journal* publish national rankings of U.S. business schools or MBA programs in addition to *U.S. News. The Economist* and *Financial Times* both publish international rankings of MBA programs. Although each ranking uses different methodologies and components, all measure graduates' salaries and/or post-graduation career development. So As such, the correlations between the most recent MBA rankings are relatively high, ranging from 0.65 to 0.85 when comparing only those schools that are listed in each ranking (excluding the *Wall Street Journal*; see Table 1). So

Global Rankings

A major development in academic quality rankings of higher education is the recent move to global comparisons. In just the past five years, two rankings have emerged that claim to list the best colleges in the world. The Shanghai Jiao Tong University's "Academic Rankings of World Universities" (ARWU), first published in 2003, pio-

neered this new frontier. Developed by the university in order to compare China's rapidly growing system of higher education to its international competitors, it has been adopted by many other nations for the same purpose, particularly after its publication in such international magazines as *The Economist*. The rankings are determined by the number of alumni and staff who have won Nobel Prizes or Fields Medals over the past 100 years (10% and 20%, respectively), the number of highly-cited researchers employed by a university (20%), the number of articles published in *Nature* and *Science* (20%), and the number of articles in the Science Citation Index-Expanded and the Social Science Citation Index (20%). A final component adjusts this research output according to the size of the institution (10%).⁵² Although these measures include labels such as "Quality of Education" and "Quality of Faculty," the ranking essentially measures a university's research prowess and has little to say about other educational measures.

The Times Higher Education of England quickly developed its own global ranking system after the appearance of ARWU. Published annually since 2005, the *Times* ranking uses a mixture of reputation, research output, and other quantitative input data to determine the top schools. Peer review surveys count for 40 percent of a school's score with another 10 percent based on surveys of employers. The remaining half of the ranking is determined by the percentage of international staff (5%) and students (5%), the number of research citations per staff member (20%) and the student-to-staff ratio (20%).53 The article accompanying the 2008 rankings noted that the top universities can generally be characterized as English-speaking (predominately American and British schools, with some from Canada and Australia) and as independent from government control (although heavily reliant on government funding). The *Times* nonetheless defends its ranking's international character by noting that the top two hundred schools are located in twenty-eight different countries.⁵⁴ Others find it easy to argue the existence of a bias in favor of schools located in the U.K. and its former colonies. 55 Regardless, the *Times* rankings receive wide coverage in foreign press. A sure indication of their continued growth in influence came in 2008 when rankings guru U.S. News and World Report published the Times results under its own banner in the United States.⁵⁶

Rankings at the global level have emerged for reasons similar to those behind their initial development at the national level. Some of the first national rankings in the U.S. were driven by graduate institutions' interest in the academic quality of programs to help them determine the intellectual strength of applicants. Today, institutions of higher education are interested in knowing the quality of foreign schools as they seek international partnerships with institutions of comparable strength. National rankings in the U.S. also became increasingly common as more and more

Americans attended college after high school, fueling the demand from consumers for information about quality. Similarly, the increasing number of students studying in foreign countries and the increasing competition to attract these students generates both a larger demand for and a greater influence of global academic quality rankings.⁵⁷ Finally, global rankings have also grown in importance as they have provided assessment and accountability for institutions of higher education, with some foreign governments going so far as to base the amount and direction of money spent on higher education on institutions' placements.⁵⁸

II. Contributions and Criticisms of College Rankings

While college rankings have been consistently subjected to serious and accurate criticisms since the furor raised in reaction against the Bureau of Education's 1911 stratification, they nonetheless serve important functions. Academic quality rankings help to make a concern for excellence in higher education publicly visible and active. The spirit of competition produced by rankings encourages universities to perform better, and combats the pitfalls of institutional stagnation that could develop in its absence. Furthermore, rankings meet a widespread demand for publicly-accessible comparative information about institutions that students will spend tens of thousands of dollars to attend. Indeed, multiple projects throughout the history of academic quality rankings have mentioned the better informing of students' college decisions as a major driving force behind the creation of their rankings, including Cattell, Prentice and Kunkel, *U.S. News*, and *Forbes*/CCAP.

James Schmotter, then assistant dean of Cornell's Johnson Graduate School of Management, argued in 1989 that colleges and universities had only themselves to blame for the rise of college rankings because higher education had failed to put forward its own system of evaluating quality that was relevant or intelligible to consumers. ⁶⁰ Two decades later, this criticism remains accurate, and college rankings continue to be an important starting place for students attempting to sort through the thousands of higher education institutions in the United States. In light of these important functions that rankings provide, Kevin Carey, a researcher for Education Sector, writes that higher education's anti-rankings sentiment is illegitimate because it reflects "an aversion to competition and accountability that ill serves students and the public at large." ⁶¹

General Criticisms

Academic quality rankings do, however, face many legitimate criticisms that should always be acknowledged and weighed in their discussion and use. One widely accepted criticism is that an ordinal ranking of colleges according to quality can produce what Gerhard Casper, then president of Stanford University, has called "false precision." Casper argued that the methodology of *U.S. News* was not precise enough to accurately identify the difference in quality between an institution ranked #1 and #2 or even #10, even though these separate ranks denote such a difference. This criticism can almost certainly be applied to all current and previous rankings. Rankings scholar Marguerite Clarke has further claimed that there may even be little difference between schools ranked #15 and #50, and found evidence that changes in an institution's rank should be viewed more as "noise" than as precise changes in quality relative to its peers. To overcome the problem of false precision, Clarke advocates listing schools alphabetically in broader "quality bands" in which each school belonging to the same band would be considered largely of equal quality.⁶²

One specific result of a highly-precise, ordinal ranking is that changes in the methodology of the ranking system can produce changes in an institution's rank without any change in the institution's quality. Minute adjustments to what criteria are used, how these criteria are weighted, and how institutions are classified and chosen for ranking can result in wide swings or even complete disappearance of a school's rank, even though the characteristics of the institution remain unchanged. In fact, one study published in May of 2008 by two mathematicians at the University of California, Berkeley, revealed that the *U.S. News* rankings are highly volatile. By adjusting the different criteria weighting and using the same data as *U.S. News*, these researchers concluded that a school's "specific placement...is essentially arbitrary." While the data which comprise the rankings are useful both for institutions and for students, the researchers argued that because weightings reflect nothing more than individual preferences, published rankings should leave criteria weighting to individual readers. 64

This research also helps to support two additional common criticisms of rankings. First, it is argued that one ranking system based on a certain set of criteria and weighting cannot possibly judge the quality of all institutions of higher education in a fair and accurate manner. Colin Diver, president of Reed College—which has refused to cooperate in providing information to *U.S. News* for over a decade—has argued that higher education is too complex of a product and that consumers' individual preferences are too diverse for all institutions to be fairly judged by a singular ranking scale. Finally, considering the lack of consensus on the exact definition of "academic quality," every ranking system makes a subjective value judgment about which criteria represent "quality" in higher education. The choice of which measures to use in a ranking implicitly and somewhat arbitrarily defines the meaning of "quality."

The above criticisms are not intended to condemn college rankings as a product. The subjectivity and at times arbitrariness of construction and adjustment of methodologies is an argument for the expansion of college rankings, not their dismissal. In light of the demand for the information provided by college rankings, multiple rankings with diverse meth-

Table 1: Correlations Between American MBA Rankings							
	Bus. Wk.	USNWR	Fin. Times	Forbes	Econ.		
Bus. Wk.	1						
USNWR	0.84	1					
Fin. Times	0.72	0.76	1				
Forbes	0.70	0.77	0.65	1			
Econ.	0.85	0.82	0.70	0.69	1		

Source: Authors' calculations.

odologies should be encouraged. This would allow consumers to evaluate the individually subjective criteria of each ranking system and adhere to the one that most closely reflects their own preferences, or to take an average of an institution's rank across many different judgments of its quality. A diversity of popularly accessible ranking systems would also help to mitigate the perverse incentives created when one ranking system is hegemonic, as discussed below.

Furthermore, the subjectivity of choosing a definition of quality does not prevent the evaluation of different ranking methodologies according to their ability to measure quality once defined. Although the choice of methodology may implicitly affect a study's definition of quality, it can be assumed that most compilers of rankings believe that quality is at the same time something independent of the criteria used to measure it. For example, most persons interpret an institution's reputation as an indication of other criteria related to their definition of quality—such as excellence of teaching, strength of academic services, etc.—rather than viewing a good reputation as itself the *definition* of quality. Once quality is subjectively defined for a given ranking system, then it is possible to make an objective evaluation of that ranking's ability to accurately measure this quality.

Reputational Rankings

With this in mind, it is useful to look at the specific strengths and weaknesses of the different approaches to academic quality ranking. Reputational surveys are one of the most widely used criteria in current ranking systems and carry some distinct benefits. Webster has argued that reputational criteria take advantage of the knowledge of "those who supposedly know most about academic quality," be they univer-

sity presidents, academic deans, department heads, or employers. Furthermore, they produce rankings that have "face validity" in that the results are consistent with the public's intuition.⁶⁶ An institution's reputation, whether among higher education officials, employers, or the general public, has been earned for a reason, after all. It seems reasonable to believe that the better an institution's reputation, the higher the institution's quality, at least at one time. Finally, a good reputation may be self-reinforcing. A school with a reputation for high quality may feel extra pressure to maintain its reputation and, therefore, have greater concern for its academic quality.

However, reputational criteria have serious drawbacks. The first is that there is evidence that an institution's reputation comes from sources that may not accurately reflect its strength in criteria that are widely accepted as good indicators of academic quality. Studies have demonstrated that an institution's reputation is closely correlated with its research productivity and its size. Publication and citation records, however, probably have little, and at best an indirect, effect on the quality of an institution's teaching and the knowledge of its graduates.⁶⁷

More importantly, even if reputation is derived from those characteristics used to define academic quality, it is highly questionable whether respondents are knowledgeable enough to judge this quality at the numerous institutions they are asked to rate. In the 1982 National Academy of Sciences ranking discussed above, respondents reported that they were, on average, unfamiliar with one-third of the programs they were asked to rate. Patricia McGuire, president of Trinity University, views as "preposterous" the idea that college presidents have the ability to rate hundreds of institutions on a scale of one to five—as they are asked to do by *U.S. News*. This lack of informed judgment by raters was revealed by the "halo effect" in Solmon and Astin's 1981 ranking of undergraduate departments, where less prestigious departments were rated more highly based on the overall reputation of their institution. In one extreme, Princeton's business department was ranked among the top ten, even though Princeton did not have an undergraduate business program.

The "halo effect" demonstrates that at least some reputational survey respondents rate programs and institutions without the requisite knowledge to make an informed judgment of quality. This information problem is of even greater concern now that global rankings with significant reputational components have emerged. A recent attempt by Germany's Center for Higher Education Development to include Switzerland's German-speaking universities in its well-regarded ranking system was hobbled by Swiss professors' considerable lack of knowledge about German schools. Considering this disconnect between countries that border each other and speak the

same language, it is highly questionable whether reputation can be used as an accurate ranking criterion on a global scale.⁷¹

Another serious concern related to raters' lack of information about the institutions they are ranking is that reputational rankings—particularly annual ones—may be creating a feedback mechanism for institutional reputations that delays the realization in the ranking of a true change in quality at an institution. Figures 1 and 2 show that the correlation between an institution's year-to-year ranking has increased to and stabilized around one hundred percent in the last decade, meaning there is little change in the rankings overall from one year to the next. Furthermore, since 1993 the reputational "peer assessment" component of the *U.S. News* rankings has increased in correlation to the overall ranking for both national universities and liberal arts colleges. In 2009, these correlations to the overall ranking were the highest of any other component for liberal arts schools and the second highest for national universities (see Tables 2 and 3).

One possible explanation of these phenomena is that raters have become more accurate in their judging of peer institutions' true quality, in that they rate institutions more in line with their overall rank as determined by the other components, and that this quality is rather stable from year to year. In light of other claims made against reputational evaluation, however, a more likely explanation may be that university presidents and admissions officials, who are asked to rate institutions about which they have little information, are turning to the most readily available source of information on academic quality: the previous year's *U.S. News* rankings. Brewer, Gates, and Goldman draw this conclusion, arguing that *U.S. News* rankings can create a positive feedback loop in which "institutions that are prestigious today are more likely to have a high level of prestige tomorrow."⁷²

Having little familiarity with even their peer institutions, academic officials are likely to allow a school's previous rank to affect their contemporary assessment of its academic reputation. Even if only indirectly, the sheer ubiquity of the *U.S. News* rankings suggests a high likelihood that previous years' rankings influence current administrators' opinions about the schools they are asked to rank. Although the institutions' characteristics of quality are likely to be relatively stable from year to year, raters relying on past rankings create a feedback loop in which one year's ranking helps to solidify their position in the next year's ranking, making changes in an institution's rank respond more slowly when actual changes in its quality do occur.⁷³

Input- and Outcome-Based Rankings

Several rankings have responded to criticisms of reputational criteria by including objective data about input variables in their calculations. The 1982 NAS study, *U.S. News*, *Money*, and *Times Higher Education* all used input variables such as library size, student-faculty ratio, incoming students' test scores, and amounts of educational expenditure either in place of or as a supplement to reputational surveys. While the strengths of these criteria include their quantitative nature and claim to objectivity, such input measures do not necessarily correlate with academic quality. Similar to a university's faculty research output, input measures are at best only indirect measures of education provided at an institution.⁷⁴

The primary alternative to reputational and input criteria is data about outcomes. Few, if any, other goods are judged by the strengths of the inputs used in their creation rather than the strength of the final output. As noted above, Prentice and Kunkel's 1930 study makes a compelling argument that institutions of higher education can best be judged by their "quality of product."⁷⁵ Indeed, as Webster has noted, all colleges seek to prepare their students for success in life after graduation; therefore, outcomes-based assessment of schools' ability to do so seems appropriate.

Unfortunately, because schools may have different ideas about the definition of success for which they are preparing students, outcomes-based rankings are especially vulnerable to the above noted criticism that no single ranking system can fully judge the complex product of higher education. A more significant criticism, however, is that outcomes-based rankings suffer from a time-lag. However post-graduation success is defined, and even if it is defined in a way that applies to all institutions of higher education, it takes time for graduates to become (or fail to become) successful. For example, it may take graduates twenty to thirty years to be listed in *Who's Who in America*, a common measurement used throughout the history of academic rankings. In that amount of time entire faculties of colleges and universities may have changed, meaning that the ranking based upon such time-lagged outcome components may have little relation to the current quality of an institution.

One final criticism of outcomes-based criteria is that they may not significantly differ from some input-based components. The research of Alexander Astin, for example, has demonstrated that a school's production of eminent alumni is largely dependent on the students' abilities when entering college. Although value-added measures that compare outcomes to inputs would mitigate this problem, no major ranking has been compiled solely or even primarily based on value-added criteria, probably because value-added data are rarely collected on a systematic basis.

III. Effects of College Rankings

These criticisms of academic quality rankings need to be carefully considered in light of the dramatic increase in their popular consumption over the past twenty-five years. Before 1983, academic quality rankings were largely compiled and used by and for institutions of higher education. These esoteric studies had begun to receive increased attention in the press in the second half of the twentieth century, but it was not until *U.S. News and World Report* began ranking schools annually in the late 1980s that an academic quality ranking was both easily understandable and circulated to millions of readers. With such exposure, the *U.S. News* rankings became *the* publicly accepted authority on academic quality, and today a school's ranking may have significant effects on the students it attracts, its admissions process, and the price of its tuition. While increased publicity carries many benefits in the form of better information for consumers, the results of these effects are arguably providing perverse incentives for institutions of higher education.

Impact of *U.S. News* on Colleges

A school's annual rank in *U.S. News* almost certainly has a relation to the caliber of students it is able to attract. A study by Patricia McDonough et al. found that when making decisions about matriculation, 59.9 percent of college freshmen listed college rankings in newsmagazines as having been "not important at all" in their matriculation decision, 29.6 percent found them "somewhat important," and only 10.5 percent found them "very important." 77 While this report has been used to downplay the effect of college rankings on matriculation decisions, such an interpretation misses the importance in the breakdown of the numbers. Indeed, the authors found that high-achieving students—those with A grades in high school and favorable assessment of their own academic ability and motivation—were highly likely to find rankings "very important." Additionally, Anne Machung has reported that parents of high-achieving high school students pay more attention to rankings than do their children, with two-thirds of them identifying U.S. News as "very helpful" in evaluating a school's quality. 78 While the majority of students may not consider a school's placement in a popular ranking,⁷⁹ those with the greatest academic ability are directly or indirectly (through their parents) influenced by rankings. This latter group is, of course, the kind of student that every college is most interested in recruiting.

Empirical evidence about the relationship between an institution's *U.S. News* rank and its incoming freshmen class further reveals the influence of rankings on high school students' college decision. Two studies have found significant effects on colleges' lagged admissions outcomes after a change in rank in *U.S. News*. The first study done by Monks and Ehrenberg looked only at selective private institutions and

found that falling one spot in the rankings results in a predicted increase of the school's rate of admittance of applicants by almost half a percentage point and a decrease in the yield rate of those students accepted. Therefore, a less prestigious rank requires a school to become less selective and admit a greater percentage of its applicant pool in order to fill its classrooms. The end result is the lowering of average SAT scores among incoming freshman. Marc Meredith's study confirmed and built on Monks and Ehrenberg's results by looking at 233 doctoral research universities, both public and private. Meredith found that a movement from the second to the first quartile increased the number of students from the top ten percent of their high school class by 1.5 percent, a 4 percent decrease in the school's admittance rate and, for public schools, an increase in average SAT scores by almost twenty points. All

Monks and Ehrenberg also explored the effects of *U.S. News* rank on a school's pricing decisions. They found that while a change in rank did not affect the (all private) schools' "sticker price," less visible discounts were associated with a decrease in rank. A less favorable rank led to lower thresholds of expected self-help and more generous financial aid grants, with a drop of ten places, resulting in a 4 percent reduction in aid-adjusted tuition and an overall decrease in net tuition. ⁸² Indeed, this ability of private schools to adjust their pricing in reaction to changes in rank was hypothesized by Meredith as explaining why the drop in SAT scores was only statistically significant for public schools. ⁸³

Perverse Incentives

Under the safe assumption that most, if not all, colleges and universities prefer to admit higher-achieving over lower-achieving students, the above effects of the *U.S. News* ranking provide strong incentives for schools to spend resources to better their position. Unfortunately, many of the components used to calculate a school's rank in *U.S. News* have a dubious connection with actual academic quality. Kevin Carey argues that 95 percent of the ranking is based directly or indirectly on institutions' "wealth, fame and exclusivity," and as a result they "focus an inordinate attention on fundraising, marketing, and attracting faculty with outsized scholarly reputations—at the expense of the core missions of access and undergraduate learning."84

While the direct impact may not be so radical as to result in higher education's "management by *U.S. News and World Report*" that Don Hossler of Indiana University decries, some of the factors in the *U.S. News* ranking surely must contribute to the inefficient use of a university's resources. For example, Hossler has provided anecdotal evidence of how the use of a school's acceptance rate in the rankings has in-

fluenced schools' admissions processes. One institution, he wrote, developed a two-part application of which the first part does not require an application fee or matter much in the final application decision, but it still counts as an application. Admission officials at the institution acknowledged that the two-tiered system was put in place in order to lower their acceptance rate and increase their overall ranking. Hossler tells of another university that now provides less information about admissions standards, hoping that more students will apply and be rejected for the same reason. While these practices probably have no impact on academic quality one way or the other, they do represent an expenditure of real resources (admissions staff's time, student applicant's time, postage, etc.) that could be saved—or put to better use—in the absence of the acceptance rate criterion.

Other components of the *U.S. News* ranking that encourage inefficiency are the measurement of expenditures on faculty resources, expenditures for research, and the peer assessment score. Thirty-five percent of the faculty resources component is determined by the average salaries of full and tenured professors. Yet, as Hossler points out, a school that increases faculty salaries or recalculates how it reports their benefits to *U.S. News* will not necessarily improve the quality of instruction, but it *will* improve the school's rank.⁸⁶ The *U.S. News* ranking also provides incentives for overinvestment in research. Spending more money on research directly rewards schools through the financial resources component, and most likely indirectly through the peer assessment component, since previous studies have demonstrated a link between reputation and research productivity. As a result, colleges and universities interested in improving their rank would have reason to push their professors to prioritize seeking grants for and spending time on expensive research projects over improving their quality of teaching, since the latter does not directly affect an institution's rank in *U.S. News*.

The use of any expenditure variable in a ranking has adverse effects on efficiency. As Ehrenberg points out, judging a school by its expenditures per students actually provides disincentives for cutting costs and keeping tuition down. In the *U.S. News* ranking, if two colleges provide the same academic quality but one does it while spending less, all other factors being equal this school would actually receive a lower ranking than the school that provided the same quality at greater cost it its students (and to taxpayers, if the school is public). In light of this, Ehrenberg notes that "no administrator in his or her right mind would take actions to cut costs unless he or she had to."⁸⁷

The concern over these perverse incentives has led some to argue that college rankings help to fuel an "academic arms race" in which colleges insatiably consume resources in hopes of increasing their prestige. However, rankings can be designed in

a way that mitigates this escalation. Regression analysis demonstrates that while there is a positive relation between an institution's instructional spending and its ranking score in U.S. News, this relationship is negative in the Forbes/CCAP ranking (see Table 4).88 Restricting the analysis to those 107 schools included in the revealed preference ranking study done by Avery et al. (see below for discussion) allows us to evaluate the relationship between spending and rank for the most academically elite schools, which are arguably the most keenly affected by the "academic arms race." The results reveal that both instructional and research spending are highly significant and positively correlated to schools' scores in the U.S. News ranking while they are highly significant and negatively correlated to schools' scores in the Forbes/CCAP ranking (see Tables 5 and 6). Neither variable is significant to a school's rank in the Avery study.

Fig. 1: Correlation Between USNWR Ranks with Previous Year's Rank (National Universities)

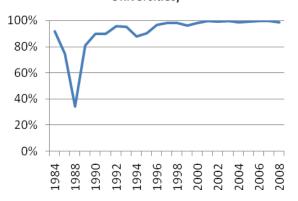
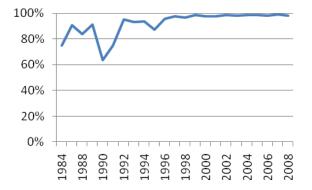


Fig. 2: Correlations Between USNWR Rank with Previous Year's Rank (Liberal Arts College)



These regressions demonstrate that while it is possible to "buy" your way up the ranks in *U.S. News*, rankings can be designed to discourage spending (*Forbes*/CCAP) or to be spending neutral (Avery et al.). These latter two ranking systems do not provide the perverse incentives for inefficient spending that exist in the *U.S. News* rankings, and instead, provide incentives for institutions of higher education to rein in their spending habits, thereby helping to disarm the "academic arms race" that has little relation to the improvement of academic quality.

IV. College Rankings Reform

In light of the criticisms and adverse effects of current ranking systems—especially those of the hegemonic *U.S. News*—the case for reform should be easily accepted. However, only reform (and not abandonment) of college rankings will result in the better use of resources in higher education. Kevin Carey argues that the rise of global rankings should completely put to rest the idea that institutions can boycott their way back to the pre-ranking days.⁸⁹ The history detailed above demonstrates that rankings have been entrenched as a part of American culture long before the

Table 2: Correlations of Component Ranks to Overall Rank in $U.S.$ News (National Universities)					
	1993	1998	2009		
Alumni Giving Rank	N/A	0.55	0.52		
Faculty Resources Rank	0.78	0.67	0.68		
Financial Resources Rank	0.62	0.66	0.77		
Graduation and Retention Rank	0.55*	0.80	0.89		
Reputation Rank	0.64	0.72	0.85		
Selectivity Rank	0.77	0.81	0.73		
Table 3: Correlations of Component Ranks to	Overall Rank in U.S. News	s (Liberal Art	s Colleges)		
	1993	1998	2009		
Alumni Giving Rank	N/A	0.47	0.68		
Faculty Resources Rank	0.64	0.48	0.65		
Financial Resources Rank	0.75**	0.65	0.79		
Graduation and Retention Rank	0.53*	0.72	0.81		
Reputation Rank	0.86	0.91	0.91		
Selectivity Rank	0.67	0.77	0.85		

first international ranking, and unequivocally supports Carey's claim that "so long as companies can publish magazines and students can choose colleges, someone will create college rankings that people will read and care about." ⁹⁰

If higher education does not want to be judged by *U.S. News* or any other newsmagazine that enters the ranking business, its only alternative is to help provide something better. As Carey writes, higher education "can't choose whether to have rankings or not, only whether they'll be good or bad." Hossler argued that data focusing on what students do *after* they enroll—including outcomes, surveys, and assessments of currently enrolled students and alumni—are more reliable indicators of a school's quality than those used in *U.S. News*. In the past, there has been a lack of better and easily usable data than that collected by *U.S. News*, but recent developments in academic quality assessment of higher education have helped to fill this void. Below are three indicators of academic quality along the lines of Hossler's prescription whose use could greatly improve college rankings and provide better incentives to institutions of higher education, along with suggestions on how best to compile such indicators into overall rankings.

The National Survey of Student Engagement

The National Survey of Student Engagement (NSSE, pronounced "Nessie") seeks to determine how successful colleges are at promoting those experiences that lead di-

rectly to student learning. The survey is administered to a representative sample of freshmen and seniors at each participating school. Categories surveyed include academic challenge, active and collaborative learning, student-faculty interaction, enriching education experiences, and supportive campus environment. Some of the specific measures within these categories include the number of books assigned and lengthy papers written, the time spent on academic and other pursuits, the synthesizing of complex ideas, how often students work with other students, how often they interact with faculty outside of class, and the availability of experiences such as study abroad and culminating senior experiences. All of the measures were chosen in light of extensive research that links them to high-quality undergraduate outcomes.⁹⁴

NSSE data show that at least some of these direct indicators of academic quality are not being captured by the U.S. News rankings. Carey reported that while some elements of the survey correlated to the magazine's rankings, a school's reputational peer review of "academic reputation"—the largest single component in U.S. News was not correlated with its promotion of active learning, student-faculty interaction or a supportive campus environment, as measured by NSSE.95 Additionally, a study by Gary Pike, after controlling for student characteristics at fourteen public research universities, found no statistically significant relation between U.S. News measures and NSSE benchmarks, except for one. Pike concluded that "the quality of a student's education...seems to have little to do with resources and reputation."96 It is not surprising, then, that two colleges that make their NSSE data public-Miles College and Jackson State University—are in the third and fourth tiers of U.S. News respectively, yet they score above the national average (sometimes dramatically so) in many of the NSSE measures. 97 Although hundreds of schools participate in the survey nationally, the most prestigious private schools largely do not. More widespread use of NSSE and a transparent reporting of the results could help improve rankings qualitatively or, in some cases, perhaps even substitute for rankings.

The Collegiate Learning Assessment and Learning Outcomes

The Collegiate Learning Assessment (CLA) is a test developed by the Council for Aid to Education (CAE) and is administered to participating schools' freshmen in the fall and seniors in the spring. Rather than testing knowledge in a multiple-choice format, CLA focuses on measuring students' abilities for critical thinking, analytic reasoning, written communication and problem solving. Tasks include interpreting, analyzing, and synthesizing information; articulating complex ideas; examining claims and evidence; and supporting ideas with relevant discussion. CAE states as a goal the simulation of complex situations that every college graduate can be expected to face some day. CLA is designed to measure the reasoning and communica-

tion skills that are largely regarded as an important outcome of a successful college education.⁹⁸

In addition to CLA's direct measuring of an institution's absolute output of students with these important skills, another major advantage of CLA is its ability to also measure the value-added element of a college. With the test administered to both freshmen and seniors, it is possible to get a picture of how the education at an institution is improving the students over four years. This helps mitigate the shortfall of other outcomes-based indicators that make it difficult to know if good outcomes are more related to the students coming into the school or to the school's practices. Furthermore, using statistical analysis, it is possible to develop a model that predicts the expected CLA scores based on the incoming students' SAT scores. This model can then reveal those schools that outperform and underperform on CLA scores compared to the SAT scores of their incoming students.⁹⁹ These value-added measurements made possible through the CLA would be much more helpful in indicating a school's academic quality than are criteria such as incoming freshmen's SAT scores. While the latter reveals the scholastic ability of students a year before entering college (and, therefore, probably tells us more about their prior education), CLA potentially reveals how much students have actually improved while at college. 100

Once again, there is evidence that the indirect measures used by *U.S. News* fail to capture the actual academic output of the colleges and universities it ranks. According to Carey, in the University of Texas system, the schools whose CLA scores most outperform the predicted results based on their incoming students' SAT scores are those that are ranked lowest by *U.S. News*. The highest ranked school, UT-Austin, falls below expected performance. ¹⁰¹ Again, CLA suffers in that it is used by only a small (but growing) minority of schools. ¹⁰²

It would also be possible to use knowledge-based tests of actual learning to measure collegiate performance. Similar to the CLA, one could test students at the beginning and end of their education in order to calculate the intervening change in knowledge to obtain a different measure of value added. Although there is one such effort that ostensibly measures "civic literacy" based on a test of knowledge of facts and principles of history, political institutions, and economics, its use has been confined to a sample of fewer than one hundred schools.

Post-Graduate Career Success

While post-graduate success of alumni has been used as an indicator of a college's quality since the earliest rankings, and as recently as the *Forbes/CCAP* 2008 rank-

ing, two recent developments in gathering improved data deserve mention as possible future reforms of college rankings. A few states have begun to link employment data about wage earnings with education data from public schools. Florida has the most advanced information system in which each student has a unique identification number that travels with him or her from K-12 and into higher education. When these records were matched up with earnings data from the state's unemployment insurance system, there were some surprising results. The four public universities in Florida that produced the highest earners were among the six worst Florida schools ranked by *U.S. News*. Although this may be partially explained by the top earners from the higher ranked schools leaving the state or going on to grad school, Carey argues that the discrepancy is too large to be completely covered by such explanations, 103 and through the use of migration data, this problem could be controlled.

Arguably the best way to judge post-graduate career success is to use employers as the rankers—or at least as the providers of data. In 2008, Boeing became the first major company to be involved in a college ranking. Using data from internal employee evaluations of its engineers, matched to data about their alma maters, Boeing created a ranking system to determine those colleges that have produced the most valuable workers in accordance with the company's goals and standards. The project found that the review of employee evaluations demonstrated significant differences in the quality of graduates from different schools, and the company has said that it will use the data to guide its decisions on hiring from and working with universities. While Boeing kept the results of the ranking confidential and shared them only with the colleges and universities themselves, the project could serve as a useful model for expanded employer-based rankings. 104

Ranking colleges using the evaluations of their graduates' employers involves significant advantages over current ranking systems. First, employers are in the best position to judge whether a graduate has been properly prepared for his or her postgraduate career, information that is vital to students who are about to make a major investment in their education. Second, expanded employer rankings that encompass many companies across a number of industries allows for the input of numerous definitions of quality. Previous rankings have often suffered from a limited definition of quality, but different employers look for different characteristics in their employers, the aggregation of which provides a picture of what schools generally best prepare their students. Finally, and related to this last point, employee evaluation-based rankings incorporate outputs that are otherwise very difficult to quantify. Other forms of outcome-based rankings, such as those measuring entries in *Who's Who* or graduates' salaries, are often criticized for having a limited definition of

"success." However, employee evaluations often combine judgments of technical skills with interpersonal skills and intangible qualities that are lost in most ranking mechanisms, yet which some schools may better develop than others.

Employer-based rankings do have their own drawbacks. First, they suffer from the same problem of all other outcomes-based rankings in that it is difficult to determine whether a highly ranked school produces good graduates because of its inherent quality or because it attracted better students to begin with because of its high rank. Second, small colleges might not be able to produce enough graduates to generate results in employer rankings that would be statistically significant. The final drawback might make a national employer-based ranking system unfeasible: a project like Boeing's carries a hefty price tag. Former acting assistant secretary for post-secondary education Cheryl Oldham acknowledged that most businesses could not spend the amount of time and money it took Boeing to complete their ranking. ¹⁰⁵ It might be possible to mitigate this cost, though, as some private entrepreneurs have developed Web sites with some postgraduate performance data by school, notably www.carreers.com.

Revealed Preference Rankings

Another potential problem with data published in both the *U.S. News* rankings and college guidebooks is its inherent susceptibility to institutional manipulation. Particular institutions can attempt to influence their rank by carefully adjusting several of the factors (especially admissions trends); there have been several well-documented cases of this phenomenon over the years. ¹⁰⁶ A study by Avery et al. sought to correct this problem by generating a "revealed preference" ranking of schools. Relying on more than just raw admissions and matriculation rates, this study ranked schools based upon student desirability. That is, a school's rank was based upon students' preference for that school relative to other schools to which they had been admitted. The study looked at the enrollment choices for high performing students by modeling the students' decisions as tournaments between competing schools. The sample of students included in the survey was not fully representative of all college applicants, and, therefore, the ranking was restricted to academically elite institutions in the United States. ¹⁰⁷

Although this ranking should not be interpreted as a direct measure of institutional academic quality, the authors argued that their methodology for the rankings has merit because college students (particularly the best students) are concerned about their peers' college choices. College students are interested in attending schools that attract the best students, because the institutional reputation derived from enrolling many good students will likely reflect favorably upon all graduates of these institu-

tions. While the authors of this rankings study did not consider their work to be a completely new and independent college ranking, they argued that the "revealed preference" of students could be an additional component used to reform current ranking systems.¹⁰⁸

Do-It-Yourself Rankings

Another recent development in college rankings is the emergence of "do-ityourself" (DIY) rankings. Rather than "pre-packaged" rankings where researchers collect data and then decide on each factor's weight in the overall ranking, DIY rankings capitalize on web technology to allow prospective students to assign their own weight to each factor so as to produce a ranking that best reflects their individual preferences. For example, after choosing a subject, the Center for Higher Education Development (DAAD) in Germany allows users to select five criteria in order of importance from an overall list of twenty-five. The resulting table lists the German universities in order of their performance on these five criteria, as well as showing the relative position of each school on each criterion. 109 In addition to their own prepackaged annual ranking, the Canadian magazine Macleans allows students to choose seven indicators (from a total of thirteen) and then assign their own numerical weight to each. The most complex DIY ranking is produced by PhDs.org, a project of former Dartmouth College professor Geoff Davis. Using data from the National Science Foundation, the National Research Council, and the National Center for Education Statistics, the website allows users to choose a subject area and then rate the importance of twenty-eight different criteria from zero to five. The detailed resulting table displays the overall rank according to the user's weighting, each department's data in the categories, and the relation of each department's data to the mean.110

Do-it-yourself rankings provide better information to consumers than typical prepackaged rankings. Two earlier mentioned weaknesses of traditional rankings are that they represent only the subjective judgments of their producers and that no one ranking system can adequately reflect what all consumers are looking for in a college. DIY rankings overcome these drawbacks by allowing consumers to receive information that is easily understood and accessible, but that also reflects their own preferences and definition of quality. The further development of DIY rankings is not a replacement for the use of better indicators of quality. Even where users can choose their own components and weights, any ranking is only as good as the data collected, and it is important that this data, whether incorporated into pre-packaged or DIY rankings, more directly and accurately reflects the quality of academic practices and outputs in higher education.

V. Conclusion

Many professors and administrators alike wistfully dream of returning to the "golden days" before college rankings allegedly ruined higher education. Such pre-ranking days, however, are now a century past. The continued development of new domestic and international ranking systems makes the chance of their demise more and more remote. The call for the abolition of ranking is not only futile, given the course of their history; it is also illegitimate in light of academic quality rankings' important contributions. Regardless of their flaws-and they have many—college rankings satisfy an immense demand for information from students, parents, and the general public about institutions that traditionally lack transparency about their internal workings and quality of product. A call for the abolition of college rankings represents a disregard for accountability that should be strongly rejected in light of the large investments that consumers and the public make in higher education.

Table 4: Dependent Variable is the Ranking Score, Ordinary Least Squares Estimation					
	Forbes/ CCAP Score	U.S News Score			
Size (FTE Enrollment dummy)	-5.4064**** (1.2618)	-4.1603* (1.7889)			
Geographic Location (NE US = 1)	-2.7316* (1.0666)	-2.5296 (1.4377)			
Endowment size per FTE Undergraduate Student	0.0000**** (0.0000)	0.0000**** (0.0000)			
Instructional Expenditures per Undergraduate	-0.0095 (0.0261)	0.0709* (0.0273)			
Percentage Receiving Fi- nancial Aid	- 20.8681**** (2.7577)	-38.1433**** (4.0434)			
Enrollment Rate	7.7364* (3.7751)	9.5231 (5.9371)			
Tuition	0.0003**** (0.0001)	0.0005**** (0.0001)			
Constant	49.9306	72.1082			
R^2	0.4387	0.6667			
N	568	256			

Note: Main entries are unstandardized regression coefficients. Numbers in parentheses are standard errors. *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001 Sources: Authors' calculations.

Like the institutions they rank, college rankings also need to be held accountable. Improvement in the design and execution of rankings requires an increased focus on accurately measuring widely agreed-upon criteria of academic quality and providing incentives for the efficient use of resources in the pursuit of this quality. Below are four guidelines for academic quality ranking systems that, if followed, will increase their usefulness and reliability. The International Ranking Expert Group (IREG), founded through a partnership between the UNESCO European Center for Higher Education and the Institute for Higher Education Policy, has already produced the

"Berlin Principles on Ranking of Higher Education of Institutions"—all of which the authors endorse. However, the following list highlights and consolidates those principles of IREG that are most relevant to the work done in this report and provides some expansion.

Define academic quality.

The preferences of potential college freshmen and the characteristics of higher education institutions are too diverse and complex for any one system of ranking to be equally useful to all. Each ranking makes a subjective judgment on what to identify as "quality." No one ranking system can legitimately claim to judge "America's Best Colleges"; it can only claim to judge those colleges that are best at the criteria it chooses to measure.112 For college rankings to adequately serve their most important function providing information to consumers about the quality of product in which they are investing—readers must be able to clearly understand a ranking's definition of "best."

Whether a ranking seeks to

Table 5: Dependent Variable is the Ranking Score, Ordinary Least Squares Estimation

5		Forbes/CCAP	U.S. News	"Revealed Preference"	
	Percent Fe- male	19.7211** (6.1873)	-7.6916 (9.0823)	-2.2496** (0.7273)	
	Percentage Receiving Fi- nancial Aid	-25.6723**** (5.9030)	-14.6678 (8.9204)	2.3670** (0.7161)	
	Instructional Expenditures per Under- graduate	-0.0469** (0.0172)	0.0669** (0.0229)	0.0031 (0.0021)	
	"Revealed Preference" Score	3.5382**** (0.7685)	3.3401** (1.0680)	N/A	
	Forbes/CCAP Score	N/A	0.4237** (0.1274)	0.0486**** (0.0106)	
	FTE Under- graduate En- rollment	-0.0003*** (0.0001)	-0.0006**** (0.0001)	0.0000 (0.0000)	
	USNWR Peer Assessment Score	10.9294**** (2.5179)	N/A	1.3954**** (0.2901)	
	Constant	9.6460	49.0529	-4.2575	
	R^2	0.7608	0.6767	0.6931	
	N	109	109	109	

Note: Main entries are unstandardized regression coefficients. Numbers in parentheses are standard errors. *p<0.05, **p<0.01, ***p<0.001, ***p<0.0001

Sources: Authors' calculations.

judge schools by the achievement level of students they attract, the educational experiences on their campuses, the strength of their professors' teaching, the success of their graduates, or some combination thereof, the ranking should make explicit what characteristics its compliers desire in colleges and universities. Only then can

consumers compare their personal preferences with those of competing ranking systems and decide which one is the best match.

<u>Use data that directly measures</u> <u>this definition of quality.</u>

Once quality has been defined, rankings should measure only those criteria that directly (or best) reflect this quality. For example, if a ranking's definition of academic quality includes the rigor of academic challenge or student-faculty interaction, the benchmarks and survey techniques of NSSE better reflect the actual level of this quality than do the components of average faculty salary or student-faculty ratio used in U.S. News. If the amount of learning that took place over four years is important to the definition of quality, then CLA much more directly measures the educational value added to an institution's graduates than do the SAT scores of incoming freshmen.

Regardless of what criteria are chosen, only those that can be related to the ranking's definition of quality

Table 6: Dependent Variable is the Ranking Score, Ordinary Least Squares Estimation

Forbes/ CCAP	U.S. News	"Revealed Preference"				
17.7724** (6.2870)	-4.8029 (9.1055)	-2.2023** (0.7392)				
-23.7215*** (6.0570)	-17.0031 (8.9019)	2.3392** (0.7310)				
-0.0616** (0.0228)	0.0867** (0.0284)	0.0024 (0.0028)				
"Revealed 3.3998**** (0.7683)	3.3159** (1.0614)	N/A				
N/A	0.4049** (0.1263)	0.0474**** (0.0107)				
-0.0004*** (0.0001)	-0.0006**** (0.0001)	0.0000 (0.0000)				
11.8895**** (2.6617)	N/A	1.4615**** (0.3116)				
5.9883	50.7491	-4.3951				
0.7605	0.6789	0.6885				
109	109	109				
	17.7724** (6.2870) -23.7215*** (6.0570) -0.0616** (0.0228) 3.3998**** (0.7683) N/A -0.0004*** (0.0001) 11.8895**** (2.6617) 5.9883 0.7605	CCAP U.S. News 17.7724** -4.8029 (6.2870) (9.1055) -23.7215*** -17.0031 (6.0570) (8.9019) -0.0616** 0.0867** (0.0228) (0.0284) 3.3159** (1.0614) N/A 0.4049** (0.1263) -0.0006**** (0.0001) (0.0001) 11.8895**** N/A 5.9883 50.7491 0.7605 0.6789				

Note: Main entries are unstandardized regression coefficients. Numbers in parentheses are standard errors. *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001 Sources: Authors' Calculations

should be considered valid. Any criterion that does not have a direct connection to the definition of quality simply adds noise to the ranking and obfuscates the true quality of the institutions judged. The relation between measure and quality should be drawn out and justified for each criterion by the rankers, thereby creating the transparency necessary for consumers to judge the validity and appropriateness of each ranking.

Use data that can only be affected by actual changes in quality.

The result of only using criteria directly related to a ranking's definition of quality is that only a true change in these criteria can produce a change in an institution's rank. For example, in rankings comprised of data from NSSE, CLA, or post-graduate success, the only way for a school to increase its position would be to actually improve its educational practices (NSSE), increase the ability of its students to think critically (CLA), or provide the skills useful in careers. In the *U.S. News* rankings—comprised mostly of input-variables—increased spending, gimmicks such as those documented by Hossler, and changes in reporting methods can all improve a school's rank even though they have questionable impact on commonplace definitions of academic quality. Many studies have come to a conclusion similar to Gary Pike's, when he wrote that academic leaders' "efforts to garner additional resources and enhance institutional reputation has little effect on the quality of their students' education." 113

A ranking based on data that can be "gamed" to improve a school's rank with little to no impact on the characteristics central to that ranking's definition of quality should not be viewed as a valid measuring tool. For this reason, using data on outcomes is usually preferable to measures of inputs. To the extent that higher education's ability to improve its students is a part of one's definition of academic quality, only the outcomes of its graduates can demonstrate an institution's relative strength. There is much less opportunity to manipulate outcomes-based data: students either perform well on the chosen measures of success or they do not. The desire for a higher rank can only be achieved by institutions improving the results of their students' education, according to the ranking's definition of success. A higher rank arrived at by spending more resources in an inputs-based ranking does not carry the same guarantee of improvement in results.

Limit perverse incentives to the extent possible.

Finally, if the criteria used in college rankings can only be affected by actual changes in quality, rankings will produce better incentives for colleges and universities to spend their resources efficiently. When an institution can only improve its rank by improving its education practices, its students' learning, or the preparedness of its graduates, it will spend more resources on these priorities. When inputs are the primary basis of a school's rank—especially expenditure variables—it will forgo these priorities and instead focus on spending money and pursuing the gimmicks mentioned above. These criteria directly encourage profligate and inefficient use of resources by rewarding schools for spending more of others' money, regardless of whether these expenditures actually contribute to academic quality.

Better criteria and design used in college rankings can help to limit the expansion of the "academic arms race" fueled by these perverse incentives. As has been shown, rankings' methodologies can be designed to reward schools for limiting their spending, while criteria such as NSSE, CLA, and post-graduate success can help to ensure that when resources are spent, they are put towards those practices that best promote educational quality. A ranking should reward—or at least not punish—a school that provides the same level of academic quality as another school but, all else being equal, does so at a lower cost to its students and the public.

As demonstrated above, improved rankings based on data such as NSSE, CLA, and post-graduate success could keep the concern for quality focused on questions of educational practices, student learning, and the preparedness of an institution's graduates, rather than on questions of prestige, money, and exclusivity. Such rankings could promote efficiency and defuse the current academic arms race by providing colleges with incentives to devote their resources toward more education-related goals. This improved data could prevent institutions from increasing their rank purely through gimmicks, profligate spending on high-profile professors and lavish student services, or policies with little relation to students' education.

However, most of the current failures of rankings published in popular news magazines to live up to these principles do not originate from willful deception or lazy methodology but from the use of second-best data. Even though NSSE and CLA are now demonstrated successes, with hundreds of schools administrating one or both, most schools only agree to participate under the condition that the results are not made public, 114 and the most prestigious schools do not participate at all. In fact, *U.S. News* does report the NSSE results for those schools that provide the data publicly, but it is far too few schools to develop a useful ranking. *USA Today* has also begun publishing NSSE summary data for public schools (presumably acquired through open-records laws) and those private schools that provide the data, but the limited number of schools and small amount of detailed data hamper the usefulness of this information. Similarly, few institutions keep complete records on their students' post-graduate success, and those who do are reluctant to release it. While there are ideas about how to collect this data through employers, they have yet to bear fruit on the necessary scale.

To a large extent, then, higher education has only itself to blame for the detrimental impacts of today's ranking systems. As Schmotter argued in 1989 at the advent of the *U.S. News* rankings, and as Carey has continued to stress today, the only way for colleges and universities to improve the means of their own assessment is to provide something better. Carey argues that the higher education establishment must "be far more transparent and forthcoming about its successes and failures than it

has historically chosen to be" if it wants a better alternative to the rankings of *U.S. News* and other popular publications. ¹¹⁵ Releasing NSSE, CLA, and post-graduate success data would be a good first step toward more transparency and would allow the producers of academic quality rankings to better serve both the consumers of and investors in colleges and universities. Higher education cannot make college rankings go away, but increased transparency and cooperation could lead to rankings that better promote true academic quality than any that have come before. Until then, it will continue to be the case that "college and university leaders who strive for high reputation- and resource-based rankings may be shortchanging their students by focusing their efforts on institutional characteristics that are largely irrelevant to a high-quality education." ¹¹⁶

The rise of do-it-yourself rankings could also have potentially revolutionary effects by offering tailor-made assessments of colleges that fit the tastes and preferences of individual consumers. The commercial aspirations of the publishers of rankings seemingly require a single set of rankings, determined by someone's judgment of what is important in evaluating university quality. But in addition to the standard one-size-fits-all evaluations, it is easy for us to see how providers increasingly may offer readers the option of custom or personalized rankings. This is particularly true if there are improvements in data sets and further advances in computer technology over time, both likely possibilities. The expansion of customizable rankings, in addition to the improvement of pre-packaged rankings, will ensure that academic quality rankings better fulfill their most important social benefit, namely, providing useful information to the consumers of higher education.

End Notes

- 1. Hossler, "College Rankings," 58. For discussion of the effects of college rankings, see below.
- 2. Webster, Academic Quality Rankings, 5.
- 3. Webster, Academic Quality Rankings, 17-18.
- 4. Ibid., 19-24.
- 5. Ibid., 20-25.
- 6. Ibid., 107-117.
- 7. Cattell, American Men, 589-592.
- 8. Webster, Academic Quality Rankings, 116-118.
- 9. Visher, American Nobles, 94-110.
- 10. Prentice and Kunkel, "Intellectual Leadership," 594.
- 11. If an entry had attended multiple schools, the person was only counted toward the school that conferred the degree and if an entry had attended a school for four years but not earned a degree, the entry was not counted.
- 12. Prentice and Kunkel, "Intellectual Leadership," 594-597; Kunkel and Prentice, "Colleges in Who's Who," 241-247.
- 13. Webster, Academic Quality Rankings, 33-34.
- 14. Kendric C. Babcock, "A Classification of Universities in the United States," in Richard Wayne Lykes, Higher Education and the United States Office of Education (1867-1953) (Washington, DC: U.S. Officer of Education), 1975, quoted in Webster, Academic Quality Rankings, 34.
- 15. Webster, Academic Quality Rankings, 38-39.
- 16. Webster, "Reputational Rankings," 238.
- 17. Bogue and Hall, Quality and Accountability, 52-53; Webster, "Reputational Rankings," 237-238.
- 18. Webster, "Reputational Rankings," 239-240.
- 19. Bogue and Hall, Quality and Accountability, 53-54; Webster, "Reputational Rankings," 244-245.
- 20. Webster, "Reputational Rankings," 245-246.
- 21. Ibid., 246-247.
- 22. A.M. Cartter. (1966). An Assessment of Quality in Graduate Education. Washington, DC: American Council on Education, quoted in Webster, "Reputational Rankings," 249.
- 23. Webster, "Reputational Rankings," 248-251.
- 24. Ibid., 249-250.
- 25. Ibid., 252.
- 26. Ibid., 253-254.
- 27. Bogue and Hall, Quality and Accountability, 56; Webster, "Reputational Rankings," 263-265.
- 28. Webster, "Reputational Rankings," 264.
- 29. Webster and Skinner, "Rating PhD Programs," 24.
- 30. Jaschik, "Baylor Pays," 2008.
- 31. Webster, "Reputational Rankings," 243-244.
- 32. Bogue and Hall, Quality and Accountability, 65; Selingo, "Self-Published College Guide."
- 33. Bogue and Hall, Quality and Accountability, 59; Webster, "Reputational Rankings," 276.
- 34. Webster, "Reputational Rankings," 276.
- 35. Webster, "Rankings of Undergraduate Education," 20.
- 36. Webster, "Reputational Rankings," 271-272; Bogue and Hall, Quality and Accountability, 60-61.
- 37. Bogue and Hall, Quality and Accountability, 62.
- 38. Webster, "Rankings of Undergraduate Education," 21-22.
- 39. Morse, "The Methodology." Whether simple graduation rates are truly a good measure of aca-

demic outcomes is open to question. Seeing as universities themselves control who they graduate and that students fail to complete their degree for myriad reasons, high graduation rates are not necessarily an indicator of high academic quality. The best outcomes-based measures are those that assess post-graduation results.

- 40. Bogue and Hall, Quality and Accountability, 63-64. Again, because of the use of graduation rates, it is unclear how accurately this model measured a school's value-added to students.
- 41. Webster, "Rankings of Undergraduate Education," 26.
- 42. See ibid., and Ehrenberg, Tuition Rising. The most recently reported incidence of questionable conduct by a higher education institution was at Baylor University, where already admitted freshmen were paid in book store credit and scholarships to retake the SAT. See Rimer, "Baylor Rewards Freshmen."
- 43. Bogue and Hall, Quality and Accountability, 64-65.
- 44. The Princeton Review, "FAQs"; The Princeton Review, "Surveying Colleges."
- 45. As research assistants at the Center for College Affordability and Productivity, both authors were involved in the development and compilation of this ranking.
- 46. Vedder and Noer, "America's Best."
- 47. Solorzano et al., "America's Best."
- 48. U.S. News & World Report, "America's Best."
- 49. Furthermore, an internet search of related terms using Google brings up very little useful information other than U.S. News's rankings, with the exception of the build-your-own graduate school rankings provided through www.PhDs.org, to be discussed below.
- 50. Business Week, "Frequently Asked Questions"; Milton, "Methodology"; Which MBA Online, "Rankings"; U.S. News & World Report, "Business Methodology"; Forbes.com, "Best Business Schools."
- 51. When comparing all schools that are listed at least once in any of the rankings, the correlation remains in essentially the same range, from 0.66 to 0.87. These correlations were performed by assigning any school not listed in a particular ranking with the rank of one higher than the highest rank of the included schools. For example, any school listed in one or more of the other four rankings, but not in the Business Week rankings, which ranks 30 schools, received a rank of 31.
- 52. Labi, "Obsession"; Carey, "Rankings Go Global"; Academic Rankings of World Universities 2007, "Rankings." For institutions that specialize in the humanities and/or social sciences, the Nature and Science component is not considered and its weight of 20% is relocated (in an undisclosed way) to the other components.
- 53. Ince, "Methodology." It is worth noting that the peer review score in 2008 was based on 6,354 surveys, a 3% response rate and the employer review on 2339 responses in 2008, response rate unreported. See QS Top Universities for a detailed discussion of the methodology.
- 54. Ince, "Ideas Without Borders."
- 55. Carey, "Rankings Go Global"; Enserink, "Who Ranks."
- 56. U.S. News & World Report, "World's Best Colleges."
- 57. See publications such as OECD's "Education at a Glance 2008" and the Institute of International Education's "Open Doors 2008" for detailed discussion of the trends in international student migration.
- 58. Labi, "Obsessions."
- 59. Enserink"Who Ranks"; Bogue and Hall, Quality and Accountability, 71-72.
- 60. Schmotter, "Colleges to Blame," 40.
- 61. Carey, "Rankings Go Global."
- 62. Clarke, "News or Noise," 47. It should be noted, of course, that when the Bureau of Education released its stratification employing the idea of "quality bands" in 1911, it was poorly received by

- higher education officials of the time. However, now that rankings are much more widespread, the return to these wider bands of quality may be welcomed by higher education.
- 63. For further discussion on how methodological changes have produced vastly different ranks for schools from one year to the next, see Ehrenberg, Tuition Rising, 60-61; Hossler, "College Rankings"; and Bogue and Hall, Quality and Accountability, 62.
- 64. Huggins and Patcher, "Selecting Universities." See also Rehmeyer, "Rating the Rankings." See below for discussion about such "do-it-yourself" rankings.
- 65. Diver, "Don't Be So Quick," 48.
- 66. Webster, Academic Quality Rankings, 142-143.
- 67. Ibid., 143; Lawrence and Green, Question of Quality, 144; Bogue and Hall, Quality and Accountability, 71.
- 68. Bogue and Hall, Quality and Accountability, 56.
- 69. Enserink, "Who Ranks."
- 70. Bogue and Hall, Quality and Accountability, 59.
- 71. Labi, "Obsessions."
- 72. Brewer, Gates, and Goldman, Pursuit of Prestige, 29-30.
- 73. The stability produced by this feedback mechanism is counter to the above claim that changes in methodology make rankings so volatile as to be unreliable. Both criticisms can be correct simultaneously, however. In the absence of any major methodological changes, the feedback mechanism may introduce too much stability from year to year.
- 74. Hossler, "College Rankings," 23.
- 75. Prentice and Kunkel, "Intellectual Leadership," 594.
- 76. Webster, Academic Quality Rankings, 118, 149.
- 77. McDonough et al., "College Rankings," 520.
- 78. Machung, "Rankings Game," 16.
- 79. It should be noted that both studies are now over ten years old. Although the empirical research has not been repeated to the authors' knowledge, it is likely that another decade of rankings hegemony for U.S. News has only increased the percentage of high school students who consider a college's ranking in their matriculation decision.
- 80. Monks and Ehrenberg, "The Impact," 6-7.
- 81. Meredith, "Universities Compete," 451-455.
- 82. Monks and Ehrenberg, "The Impact," 7-8.
- 83. Meredith, "Universities Compete," 456.
- 84. Carey, "Beat Them," 50.
- 85. Hossler, "College Rankings," 21-22.
- 86. Ibid., 24.
- 87. Ehrenberg, Tuition Rising, 16, 58.
- 88. Such a result is not surprising, as U.S. News directly and positively measures instructional spending per student, while the Forbes/CCAP ranking includes a component that rewards schools for keeping low the level of debt their students accrue while attending.
- 89. Carey, "Rankings Go Global."
- 90. Ibid.
- 91. Ibid.; Carey, "Beat Them," 52.
- 92. Hossler, "College Rankings," 23.
- 93. Carey, "Beat Them," 52.
- 94. National Survey of Student Engagement, "Our Origins"; Carey, College Rankings Reformed, 4.
- 95. Carey, College Rankings Reformed, 5.
- 96. Pike, 204.

- 97. Carey, College Rankings Reformed, 5.
- 98. Council for Aid to Education, "Collegiate Learning."
- 99. Carey, College Rankings Reformed, 7.
- 100. Carey, "Beat Them," 52.
- 101. Carey, College Rankings Reformed, 8.
- 102. Intercollegiate Studies Institute, "Our Fading Heritage."
- 103. Ibid., 12.
- 104. Basken, "Boeing to Rank Colleges."
- 105. Ibid.
- 106. Jaschik ("Baylor Pays") reveals such a case. In 2008 Baylor University offered financial aid incentives to admitted freshmen who retook the SAT and received higher score.
- 107. Avery et al., "A Revealed Preference."
- 108. Ibid.
- 109. Center for Higher Education Development, "CHE University Ranking."
- 110. PhDs.org, "Find the School."
- 111. Traditional rankings will likely maintain a strong role when it comes to judging the academic quality of higher education because many high school students may not be clear on what factors will be most important to them ex ante. Therefore, students and parents will be likely to continue to find "expert" judgments of a school's quality helpful. However, it is desirable that producers of pre-packaged rankings provide the ability to build individualized rankings using the same data in the future.
- 112. Pike, 206.
- 113. Ibid., 204. See Hossler, "College Rankings"; Carey, "College Rankings Reformed"; and Ehrenberg, Tuition Rising for additional discussion about the relationship between input-variables and true academic quality.
- 114. Carey, College Rankings Reformed, 4.
- 115. Carey, "Beat Them," 52.
- 116. Pike, 206.

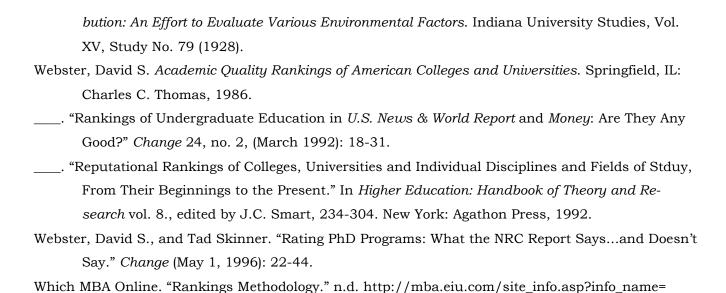
Bibliography

- Academic Rankings of World Universities 2007. "Rankings Methodology." http://ed.sjtu.edu.cn/rank/2007/ARWU2007Methodology.htm (accessed December 18, 2008).
- Avery, Christopher, Mark Glickman, Caroline Hoxby, and Andrew Metrick. "A Revealed Preference Ranking of U.S. Colleges and Universities." National Bureau of Economic Research, Working Paper 10803, September 2004.
- Basken, Paul. "Boeing to Rank Colleges by Measuring Graduates' Job Success," *The Chronicle of Higher Education*, September 19, 2008. http://chronicle.com/free/v55/i04/04a00102.htm (accessed December 31, 2008).
- Bogue, E. Grady, and Kimberely Bingham Hall. *Quality and Accountability in Higher Education: Improving Policy, Enhancing Performance.* Westport, CT: Praeger Publishers, 2003.
- Brewer, Dominic J., Susan M. Gates, and Charles A. Goldman. *In Pursuit of Prestige: Strategy and Competition in U.S. Higher Education*. New Brunswick: Transaction Publishers, 2002.
- Business Week. "Frequently Asked Questions: MBA Programs." November 13, 2008. http://www.businessweek.com/bschools/content/nov2008/bs20081113_320726_page_2.htm (accessed December 30, 2008).
- Carey, Kevin. "Beat Them, Join Them, or Leave Them Be?" *Currents* 33, no. 6, (July/August 2007): 50-52.
- _____. College Rankings Reformed: The Case for a New Order in Higher Education. Education Sector Reports, September 2006.
- ____. "Rankings Go Global." *Inside Higher Ed*, 6 May 2008. http://www.insidehighered.com/views/2008/05/06/carey (accessed December 19, 2008)
- Cattell, James McKeen. American Men of Science: A Biographical Dictionary. New York: Science Press, 1910.
- Center for Higher Education Development. "CHE University Ranking." http://www.daad.de/deutschland/hochschulen/hochschulranking/06543.en.html (accessed December 9, 2008).
- Clarke, Marguerite. "News or Noise? An Analysis of *U.S. News and World Report's* Ranking Scores." *Educational Measurement: Issues and Practice* 21, no. 4 (Winter 2002): 39-48.
- Council for Aid to Education. "Collegiate Learning Assessment (CLA)." n.d. http://www.cae.org/content/pro_collegiate.htm (accessed December 9, 2008).
- Diver, Colin S. "Don't Be So Quick to Join the Refusenik Ranks." *Currents* 33, no. 6 (July/August 2007): 48-49.
- Ehrenberg, Ronald G. *Tuition Rising: Why College Costs So Much.* Cambridge, MA: Harvard University Press, 2000.
- Enserink, Martin. "Who Ranks the University Rankers?" *Science* 317, no. 5841 (August 24, 2007): 1026-1028. http://www.sciencemag.org/cgi/content/full/317/5841/1026 (accessed December 18, 2008).

- Forbes.com. "Best Business Schools: Methodology." August 16, 2007. http://www.forbes.com/2007/08/16 /business-school-methodology-biz-07mba_cz_kb_0816bschoolmeth.html (accessed December 30, 2008).
- Hossler, Don. "The Problem with College Rankings." About Campus (March/April 2000): 20-24.
- Huggins, Peter, and Lior Pachter. "Selecting Universities: Personal Preference and Rankings." Working Paper, May 7, 2008, unpublished.
- Ince, Martin. "Ideas without Borders as Excellence Goes Global." *Times Higher Education*, November 8, 2007. http://www.timeshighereducation.co.uk/story.asp?storycode=311410 (accessed December 18, 2008).
- Intercollegiate Studies Institute. "Our Fading Heritage: Americans Fail a Basic Test on Their History and Institutions." 2008-2009. http://www.americancivicliteracy.org/ (accessed December 9, 2008).
- International Ranking Expert Group (IREG). "Berlin Principles on Ranking of Higher Education Institutions." May 20, 2006. http://www.ihep.org/assets/files/publications/a-f/BerlinPrinciplesRanking.pdf (accessed December 9, 2008).
- Jaschik, Scott. "Baylor Pays for SAT Gains." *Inside Higher Ed*, 15 October 2008. http://www.insidehighered.com/news/2008/10/15/baylor (accessed December 9, 2008).
- Kunkel, Beverly W., and Donald B. Prentice. "The Colleges in Who's Who in America." *School and Society* 74, no. 1922 (October 20, 1951): 241-247.
- Labi, Aisha. "Obsession With Rankings Goes Global." *The Chronicle of Higher Education*, October 17, 2008. http://chronicle.com/weekly/v55/i08/08a02701.htm (accessed December 18, 2008).
- Lawrence, Judith K., and Kenneth C. Green. *A Question of Quality: The Higher Education Ratings Game.* AAHE-ERIC/Higher Education Research Report No. 5, 1980.
- Machung, Anne. "Playing the Rankings Game." Change 30, no. 4 (July/August 1998):12-16.
- McDonough, Patricia M., Anthony Lising Antonio, Mary Beth Walpole, and Leonor Xóchitl Pérez.

 "College Rankings: Democratized Knowledge for Whom?" *Research in Higher Education* 39, no. 5 (1998): 513-537.
- Meredith, Marc. "Why Do Universities Compete in the Rankings Game?" *Research in Higher Education* 45, no. 5 (August 2004): 443-461.
- Milton, Ursula. "Methodology: How the FT Compiles Its MBA League Tables." Financial Times, January 28, 2008. http://www.ft.com/cms/s/2/3e262cfe-cb17-11dc-97ff000077b07658,dwp_uuid= 06879246-a4b2-11dc-a93b-0000779fd2ac.html (accessed December 30, 2008).
- Monks, James, and Ronald G. Ehrenberg. "The Impact of *U.S. News & World Report* College Rankings on Admissions Outcomes and Pricing Policies at Selective Private Institutions." Cambridge, MA: National Bureau of Economic Research, Working Paper 7227, July 1999.

- Morse, Robert J. "The Methodology." U.S. News & World Report, 18 September 1995.
- National Survey of Student Engagement. "Our Origins and Potential." n.d. http://nsse.iub.edu/html/origins.cfm (accessed December 9, 2008).
- PhDs.org. "Find the Graduate School That's Right for You." n.d. http://graduate-school.phds.org/ (accessed December 9, 2008).
- Pike, Gary. "Measuring Quality: A Comparison of *U.S. News*Rankings and NSSE Benchmarks." *Research in Higher Education* 45, no. 2 (March 2004): 193-208. Prentice, Donald B., and Beverly Waugh Kunkel. "The Colleges' Contributions to Intellectual Leadership." *School and Society* 32, no. 827 (November 1, 1930): 594-600.
- The Princeton Review. "FAQs: The Best 368 Colleges." n.d. http://www.princetonreview.com/faqs-college-rankings.aspx#Q1 (accessed December 12, 2008).
- ____. "Surveying Colleges: How We Do It." n.d. http://www.princetonreview.com/how-we-do-it.aspx (accessed December 12, 2008).
- QS Top Universities. "Background/Methodology." n.d. http://www.topuniversities.com/worlduniversity rankings/methodology/peer_review/ (accessed December 18, 2008).
- Rehmeyer, Julie. "Rating the Rankings." *Science News*, October 3, 2008. http://www.sciencenews.org/view/generic/id/37207/title/ Math_Trek__Rating_the_rankings (accessed December 9, 2008).
- Rimer, Sara. "Baylor Rewards Freshmen Who Retake SAT." *The New York Times*, 14 October 2008. http://www.nytimes.com/2008/10/15/education/15baylor.html?_r=1 (accessed December 9, 2008).
- Schmotter, James W. "Colleges Have Themselves to Blame for the Influence of Journalistic Rankings of Their Quality." *The Chronicle of Higher Education*, August 16, 1989, A40.
- Selingo, Jeffrey. "A Self-Published College Guide Goes Big-Time, and Educators Cry Foul." *The Chronicle of Higher Education*, November 7, 1997. http://chronicle.com/free/v44/i11/11a00101.htm (accessed December 9, 2008).
- Solorzano, Lucia, et al. "America's Best Professional Schools: Top Training for Top Jobs." *U.S. News* & *World Report*, 2 November 1987, 70.
- U.S. News & World Report. "America's Best Graduate Schools 2009." n.d. http://grad-schools.usnews.rankingsandreviews.com/grad (accessed December 30, 2008).
- _____. "Business Methodology." March 26, 2008. http://www.usnews.com/articles/ education/best-graduate-schools/2008/03/26/business-methodology.html (accessed December 30, 2008).
- Vedder, Richard and Michael Noer. "America's Best Colleges." August 13, 2008. http://www.forbes.com/2008/08/13/college-university-rankings-oped-college08-cx_rv_mn_0813intro.html (accessed December 19, 2008).
- Visher, Stephen. Geography of American Notables: A Statistical Study of Birthplaces, Training, Distri-



wmba rank method (accessed December 30, 2008).