College and University Ranking Systems

GLOBAL PERSPECTIVES AND AMERICAN CHALLENGES

INSTITUTE FOR HIGHER EDUCATION POLICY (EDITOR) • APRIL 2007

The Institute for Higher Education Policy (IHEP) is an independent, nonprofit organization that is dedicated to access and success in postsecondary education around the world. Established in 1993, the Washington, D.C.-based organization uses unique research and innovative programs to inform key decision makers who shape public policy and support economic and social development. The IHEP's Web site, www.ihep.org, features an expansive collection of higher education information available free of charge and provides access to some of the most respected professionals in the fields of public policy and research.

Institute for Higher Education Policy
1320 19th Street, NW, Suite 400, Washington, DC 20036
Phone: (202) 861-8223 • Facsimile: (202) 861-9307 • Internet: www.ihep.org

College and University Ranking Systems

GLOBAL PERSPECTIVES AND AMERICAN CHALLENGES

INSTITUTE FOR HIGHER EDUCATION POLICY (EDITOR) • APRIL 2007

Acknowledgments

The Institute for Higher Education Policy (IHEP) is grateful to chapter authors Alvin P. Sanoff, Alex Usher, Massimo Savino, and Marguerite Clarke for their efforts in making this monograph a success. In addition, IHEP staff members Jamie P. Merisotis, Alisa F. Cunningham, Arnold M. Kee, Tia T. Gordon, and Lacey H. Leegwater edited the monograph and contributed to its introduction and "next steps" sections. We would also like to thank Jan Sadlak and the UNESCO European Centre for Higher Education for their continued partnership with IHEP on rankings work, and members of the International Rankings Expert Group for their contributions to the international rankings dialogue informing this work.

IHEP would also like to thank Lumina Foundation for Education for its generous funding of this project. Lumina Foundation for Education (www.luminafoundation.org), an Indianapolis-based, private, independent foundation, strives to help people achieve their potential by expanding access and success in education beyond high school. Through grants for research, innovation, communication, and evaluation, as well as policy education and leadership development, Lumina Foundation addresses issues that affect access and educational attainment among all students, particularly underserved student groups, including adult learners. The Foundation bases its mission on the belief that postsecondary education remains one of the most beneficial investments that individuals can make in themselves and that society can make in its people.

The opinions expressed in this report are those of the author(s) and do not necessarily reflect the views of Lumina Foundation for Education. In addition, the chapters of this report were commissioned by IHEP and reflect the views of the chapter authors.

Table of Contents

Introduction

When *U.S. News & World Report* began its ranking of American colleges in 1983, publishers in other countries quickly followed with their own hierarchical measures, providing consumer information (and opportunities for institutional marketing) while attempting to impact the quality of higher education. In the course of the last two decades, higher education ranking systems and "league tables" (as they are referred to in the United Kingdom and elsewhere) have emerged in dozens of countries.[1] These rankings are conducted not only by media in the private sector, but also by professional associations and governments.

Over the decades since higher education rankings first appeared, numerous debates have surfaced about their methodologies, objectivity, impact on colleges and universities, and role in the structure of accountability within nations that use them. In recent years, as many countries have introduced tuition fees, and as tuition prices have escalated in the United States and elsewhere, rankings have been the focus of increased scrutiny. Although there has been significant research, especially in the United States, about the ways in which rankings might be improved, there has been less research on what other countries have been doing and how their ranking systems differ from U.S. rankings. In addition, there has been very little research on how rankings may impact students' access to postsecondary education, their selection of particular colleges, and their paths to graduate from school and/or find employment.

Our goal for this monograph is to better understand the ways in which ranking systems function and how lessons learned from other countries that use higher education ranking systems might influence similar practices in the United States. Toward this end, this monograph chronicles recent efforts that have brought together rankers and researchers from around the world to study higher education rankings. The monograph includes three papers that were commissioned to examine various perspectives on rankings around the world and lessons they might provide for rankings in the United States.

Background on Rankings and League Tables

Ranking approaches and systems, like higher education institutions, vary extensively and are often tied to the unique higher education context of a given nation. In general, however, each system or approach tends to include a similar, logical set of elements. First, data are collected, either from existing sources or from original surveys. Following this, the type and quantity of variables are selected from the information gathered. Next, the indicators are standardized and weighted

[1]In general, when this report refers to "rankings," we mean both rankings in the *U.S. News* sense of the term *and* what are known in Britain and elsewhere as league tables. We recognize, however, that the two terms are not precisely interchangeable; some of the differences are discussed in this monograph in the chapter "A Global Survey of Rankings and League Tables," by Alex Usher and Massimo Savino.

from the selected variables. Finally, calculations are conducted and comparisons made so that institutions are sorted into "ranked order."

Higher education rankings are often controversial and heavily debated in local, national, and, increasingly, international contexts. Whether or not colleges and universities agree with the various ranking systems and league table findings, however, ranking systems clearly are here to stay. Assuming that ranking systems will persist, it follows that we would want to determine how best they might be constructed. In other words, what types of performance indicators, procedures, and ethical considerations should be included in a conceptual framework or typology for higher education ranking systems?

Current methodologies exhibit various strengths and weaknesses. Different rankings include indicators that students may overlook when considering an institution's quality. These rankings allow institutions to distinguish themselves based on who they are and what they do for consumers of higher education. The competition sparked by rankings methodologies also has both strengths and weaknesses. Some will argue that competition indirectly improves overall quality in the higher education market. Others may argue that the same competitive forces skew institutional policies in ways that might cause college or university personnel to work against their own missions.

Nevertheless, the inherent weaknesses of rankings methodologies often overshadow their strengths. In fact, the major flaw in rankings may be their continual changes in methodology. For instance, although institutions may not actually change in a significant way, ratings can fluctuate year-to-year as rankers change the weights assigned to different indicators. Likewise, many ranking systems produce a single number that summarizes the overall ranking of an academic institution. This practice makes it difficult for students to distinguish among institutions based on the characteristics they find most important. Additionally, much of the objective data used in the rankings are self-reported by the institutions. Continuing such a practice without external validation of data could lead to difficulties for rankings in the future, especially if institutions continue to perceive that rankings influence consumer behavior.

Recent Efforts to Examine the Issues

The Institute for Higher Education Policy (IHEP) has been involved in the global dialogue about higher education rankings since 2002. IHEP has maintained its involvement in these conversations for a variety of reasons, including the potential links among rankings, institutional policies, and opportunities for underrepresented students to succeed in higher education.

In 2002, IHEP President Jamie P. Merisotis served as rapporteur at a ground-breaking international meeting convened to examine the "functioning" of higher education ranking systems and league tables. The meeting, held in Warsaw, Poland, and sponsored by the UNESCO European Centre for Higher Education (UNESCO-CEPES, headquartered in Bucharest, Romania), featured papers and presentations from, among other countries, Japan, Germany, Nigeria, Poland, the Russian Federation, the United Kingdom, and the United States. Some 40 participants from 12 countries, representing journals that regularly publish rankings of higher education institutions and including top-level experts from national bodies and international governmental and non-governmental organizations, discussed various issues related to rankings. One key outcome of the

meeting was that further work is needed to improve the conceptual frameworks, methodologies, and organizational aspects of college rankings.

A follow-up meeting, held in December 2004 in Washington, D.C., was hosted jointly by UNESCO-CEPES and IHEP. That meeting included more than 20 leading experts from around the world who either conduct rankings or analyze those ranking systems. One outcome of that meeting was the establishment of an International Rankings Expert Group (IREG), composed largely of the participants in the Washington meeting.

IREG held a third international meeting on rankings, in Berlin, Germany, in May 2006. That meeting was organized by the Centre for Higher Education Development (*Centrum für Hochschulentwicklung*) in Germany, UNESCO-CEPES, and IHEP.

At the Berlin meeting, IREG participants—including representatives who work on the rankings published by *U.S. News & World Report,* the *Times Higher Education Supplement* in London, *Die Zeit* in Germany, *Asahi Shimbun* in Japan, and leading thinkers from Russia, China, the Netherlands, and other nations—met to discuss how ranking system methodologies might be enhanced in order to provide better and more detailed information to consumers.

Participants in the Berlin meeting discussed ranking methodologies in various countries and lessons learned from the experiences of those countries. The meeting included an in-depth discussion of how ranking methodologies might be improved. In addition, participants discussed the fact that many ranking systems classify only a small percentage of institutions and fail to capture the diversity of institutions. Participants emphasized the differences among rankings, as well as the relationship between ranking systems and national goals in the context of an increasingly competitive global economy. (It was clear, for example, that many countries want to develop world class universities that will allow them to attract top faculty and contribute to their economic growth.) There was some agreement that privatization has increasingly turned higher education into a consumer good.

The participants described controversies in their home countries when rankings were developed, especially reaction by colleges and universities and feelings that such a system is an alien concept. It was also noted that in some cases rankings have been useful to national governments in reforming higher education systems and promoting a culture of quality and transparency. In addition, participants observed, rankings may be useful to the colleges themselves in terms of benchmarking and strategic planning, and may be useful to students when there is a lack of publicly established criteria for quality. Nonetheless, participants noted that reactions to rankings depend on the context of individual countries, and may have unintended impacts on equality, institutional behavior, governance, and other issues.

An important outcome of the Berlin meeting was the development of a framework for the elaboration and dissemination of rankings that ultimately will lead to a system of continuous improvement and refinement of rankings systems. The Berlin Principles for good ranking practices (included as an appendix in this monograph) will be useful for the improvement and evaluation of rankings around the world as rankings practitioners continue to refine their methodologies.

The Berlin Principles include a number of recommendations for what rankings should and should not do, within a number of thematic areas: purposes and goals; design and weighting of indicators; collection and processing of data; and the presentation of rankings results. Some important suggestions include the following:

- Rankings can provide comparative information and improved understanding of higher education, but should not be the main method for assessing what higher education is and does; rather, they can complement the work of government, accrediting authorities, and independent review agencies.
- Rankings must recognize the diversity of institutions and take the different missions and goals of institutions into account. Quality measures for research-oriented institutions, for example, are quite different from those that are appropriate for institutions that provide broad access to underserved communities.
- The choice of methods used to prepare rankings should be clear and unambiguous. This transparency should include the calculation of indicators as well as the origin of data.
- The choice of data should be grounded in recognition of the ability of each measure to represent quality and academic and institutional strengths, and not availability of data. In addition, rankings should measure outcomes in preference to inputs whenever possible. Measures of outcomes provide a more accurate assessment of the standing and/or quality of a given institution or program.
- If weights are assigned to different indicators, rankings should note that prominently. Moreover, there should be only limited changes to such weightings. Changes in weights make it difficult for consumers to discern whether an institution's or program's status changed in the rankings due to an inherent difference or due to a methodological change. In addition, rankings should provide consumers with a clear understanding of all of the factors used to develop a ranking, and offer consumers a choice in how rankings are displayed.
- And finally, rankings should apply measures of quality assurance to ranking processes themselves. These processes should take note of the expertise that is being applied to evaluate institutions and use this knowledge to evaluate the ranking itself. Rankings should be learning systems that continuously apply this expertise to develop methodology.

These principles address some of the ways in which rankings could be improved in order to better serve the needs of consumers, institutions, and governments. The principles should be kept in mind when examining the role and impact of rankings within the United States.

Themes of this Monograph

To initiate discussion at the Berlin meeting, IHEP, in consultation with UNESCO-CEPES, commissioned three papers by leading independent analysts of higher education rankings from around the world. The authors presented their papers at the meeting. The papers were grouped around several themes:

- The history and development of a major ranking system used in a specific nation (in this case, *U.S. News & World Report* in the United States), including the issues and challenges faced. Did the approach of the methodology achieve its purpose?

- A framework for understanding ranking systems globally, including the key elements and types of rankings that exist around the world and how they are constructed. How do rankings differ among countries? Why are rankings growing in prominence and importance?
- The impact of rankings on higher education access, choice, and opportunity, especially for disadvantaged students. Have rankings contributed to a growing stratification within higher education systems?

These themes have implications for U.S. systems of rankings. The lessons learned from the global examples will shape the domestic debate and will likely then be applied to improve the different ranking methodologies used in the United States. In turn, the experiences of the United States as it changes and adapts to policy development may re-emerge to influence the global debate.

Summary of the Findings

The Development of the *U.S. News* Rankings

In the first chapter of this monograph, Alvin Sanoff, who served as managing editor of the rankings project at *U.S. News & World Report* for many years, outlines the history of the *U.S. News* rankings, including the succession of changes that have been made as a result of criticisms of the magazine's rankings methodologies.

According to Sanoff, the *U.S. News* rankings historically were a marketing device for the magazine. As the rankings grew to include more institutions, they also became more prominent in the media. Public demand for the rankings increased as the price of college grew and consumers wanted ways to differentiate among institutions.

The emerging prominence of the *U.S. News* rankings resulted in a series of critiques and responses. An early criticism, for example, that the rankings should use a multi-dimensional approach, led to the addition or alteration of categories. Because critics noted that institutional "scores" gave the illusion of precision when many colleges were actually separated by miniscule amounts, the magazine began to round scores. When some colleges began to toy with data they submitted in order to increase their ranking, the magazine responded by tightening their data definitions and cross-checking it with other sources.

Sanoff notes that as a result of these criticisms, the magazine has made its methodology more transparent and, with the help of dialogue with institutions, guidance counselors, and others, has modified its methodology over the years to try to improve it. To this end, the magazine has also met frequently with college officials and other stakeholders.

One lesson for the United States to draw from this chapter is that—as noted in the Berlin criteria—it is important for ranking systems to reevaluate their methodologies periodically and note how rankings are perceived by their various audiences. In the case of *U.S. News,* even though the percentage of students attending the top institutions that the magazine ranks may be relatively small compared to the entire student population, there appears to have been a trickle-down effect into the rest of the U.S. higher education system, which has encouraged competition and reinforced the use of rankings as a perceived measure of quality. This increasingly competitive environment makes self-evaluation even more important. *U.S. News* has periodically updated its methodologies and

goals, often under pressure from institutions and consumers, and it is important to continue this pressure as the demand for institutional accountability increases. At the same time, it is important to recognize that most ranking systems are promulgated by the media, and their goals may conflict with national or state goals for accountability and postsecondary education opportunity.

Understanding Rankings Around the World

In the second chapter, Alex Usher and Massimo Savino review a number of rankings systems and league tables worldwide. These include the traditional national systems that rank colleges within a country against each other, as well as the new variation of rankings that rank colleges across national borders. All of these systems compare institutions across a range of indicators, in a manner similar to that used with performance indicators. In the case of league tables, the indicators are then turned into a "score" using a specific weighting scheme.

Usher and Savino explain the variation in data sources used to rank institutions, and the positives and negatives associated with each type of data—survey data; institutional data; and independent, third-party data (such as those collected by governments). The authors examine in depth the types of indicators used by ranking systems across the globe, and how indicators differ among countries yet are often similar within specific regions or cultures. They also explain how the combination of indicators used in each ranking system reflects the ranker's view of "quality" for higher education institutions (and/or the lack of data to measure other indicators). Finally, they offer an example of a ranking system that does not use an overall score, but rather collects indicator data from individual departments, presents all of the indicators on a Web site, and allows consumers to choose their own weighting scheme.

Given the impact on the structure of higher education systems throughout the world, it is essential to think about rankings within the context of national goals.

Overall, this chapter demonstrates that rankings are increasingly being used as a measure of quality, which may be defined in different ways and may be measured by a variety of indicators, depending on the perspective of a ranking's creators. Thus, the goals of rankings systems may differ by region, by the higher education system's stage of development, and by the entity conducting the ranking. Given the impact on the structure of higher education systems throughout the world, as well as their role as a tool of accountability, it is essential to think about rankings within the context of national goals—should the rankings' visions of quality become the de facto measure of accountability in the eyes of consumers as well as governments? Consumers and governments may define quality or accountability differently. This potential dissonance should be a part of the public debate, leading to agreement on valid measures of success. This potential agreement could then affect decisions on the kinds of data colleges and governments should collect on an annual basis to substantiate claims of success.

The Impact of Rankings on Higher Education Access, Choice, and Opportunities

In the third chapter, Marguerite Clarke examines the impact of rankings on student access to higher education, choice of institution, and opportunities after graduation, especially for students from disadvantaged backgrounds. Her findings show that access may be threatened by rankings in certain circumstances—by creating incentives, for example, for schools to recruit students who will maintain or enhance their position in the rankings. Such practices in the United States contribute to the stratification of the higher education system and, in turn, encourage such institutional

policies as early admission decisions, merit aid, and tuition discounting. Clarke finds that these consequences, along with the increasing expenditures to meet consumer demands for dorms, technology, and facilities, converge to the detriment of disadvantaged students. She also notes that similar stratification related to rankings is occurring in other countries with diversified higher education systems.

Clarke finds less evidence of an impact of rankings on college choice among students who do enroll in college. In general, students with higher income and/or high-achieving students are the most likely to use rankings to help guide college choice. (This pattern holds both for the United States and for other countries with available data.) This finding is not surprising, given that disadvantaged students are more likely to attend less selective institutions and community colleges, which generally are not ranked by *U.S. News* or other organizations.

When considering students' opportunities after graduation, Clarke finds that in the United States the perceived status of the degree-granting institution (which, as Clarke points out, relates strongly to rankings) influences a student's employment opportunities and earnings. Some of the research she cites suggests that any impact for most students may be temporary, while other research indicates there is an impact for professional students and low-income students. (Of course, low-income students are less likely to attend highly selective institutions in the first place.) Clarke notes that only a small amount of research (mainly U.S. related) directly examines the effects of rankings on employment and earnings outcomes; however, these data suggest that rankings do have an impact on outcomes in these areas, at least for business school graduates.

Clarke found it difficult in most cases to compare the outcomes of rankings within or across countries, especially developing nations, due to a lack of data. However, the introduction of global ranking systems is likely to provide a better means of comparison in the future. These developments mean more data for researchers. Also, anecdotal evidence suggests that new ranking systems being developed within and among countries may better inform prospective students about the types of institutions that are available for them to attend. Cautions remain, however, given the high likelihood that the introduction of ranking systems will encourage higher education stratification—and, in fact, some countries are explicitly moving in that direction. Certain countries see rankings and subsequent stratification as the means to create "world class" universities and thus meet increasing global competition.

One lesson for the United States is to place a policy emphasis on addressing the stratification of students by socioeconomic status and develop new and innovative solutions to the problem. For example, in the last year, several institutions in the highest tier of the rankings have eliminated their early admissions programs with the hope of serving more low-income students. (Several of these schools have also reduced or effectively eliminated tuition and other expenses for low-income students.) It is unclear whether this will impact the policies of other institutions, or modify the enrollment patterns of students as a whole. However, it is clear that the influence of rankings on postsecondary opportunity, whether direct or indirect, must be part of the broader debate on whether a more market-based system of higher education in the United States is changing institutional behavior in desirable ways.

In addition, although these chapters focus most of their attention on the impact of rankings in the United States, the argument about how stratification may impact access and opportunity may provide a lesson for developing nations as they strive to emulate or compete with the United

States. As developing countries continue efforts to enroll increasing numbers of students, making higher education more accessible to more students, while at the same time enrolling higher numbers of disadvantaged or underrepresented students, they may find a conflict between the goal of making education accessible and the use of stratification of higher education to help create "world class" universities.

Ongoing Challenges

Taken together, these chapters describe challenges faced by countries around the world in implementing rankings or improving them. In coordination with the work of the IREG meetings, the findings can also contribute to the continued improvement of rankings in the United States by clarifying their role and holding a mirror to once-mysterious methodologies.

The following chapters substantiate the need for a process of ongoing revision and greater transparency in ranking systems, especially as those systems become part of the accountability structure. If this trend continues, improvements in the collection of data will also be necessary. Moreover, the collected data should be stored in the public domain so that researchers and consumers can sift through it and test their validity. Such a public system would allow more research questions to be examined and would help inform the global dialogue about how to use rankings as an innovative tool for reflecting the market's perspective on higher education quality.

The *U.S. News* College Rankings: A View from the Inside

By Alvin P. Sanoff

U.S. News & World Report was the pioneer in ranking colleges and universities, and 2006 marked the twentieth anniversary of its annual rankings. What *U.S. News* started has since gone global. The magazine paved the way for what has become a worldwide phenomenon.

This chapter will endeavor to provide a brief history of the magazine's college rankings, discuss their evolution, highlight some key changes made over the course of two decades, and talk about the factors both within and outside of higher education that help explain the project's success. It will touch on the magazine's graduate school rankings. But the focus is on *U.S. News's* ranking of undergraduate institutions.

As the managing editor of the rankings project for almost seven years, from 1992 to 1998, I was part of the decision-making team and helped handle a crisis that threatened the viability of the project. This paper, then, is written from an insider's perspective.[2]

In the Beginning

The rankings actually began in 1983, but they did not become an annual event until 1987. They were launched with little fanfare. Editors thought the project was worth experimenting with because it might garner attention and sell magazines. No one imagined that the rankings would become what some consider the 800-pound gorilla of American higher education, important enough to be the subject of doctoral dissertations, academic papers and conferences, and endless debate.

In its initial foray into rankings, *U.S. News* took a simple approach. It surveyed college presidents, asking them to identify the nation's best institutions of higher learning. The 1983 survey was an outgrowth of an unrelated project that had become a staple of the magazine: a survey of U.S. leaders to identify the most influential Americans. That annual project received a lot of attention. Editors thought they could try a somewhat similar approach to identify the best colleges.

The college survey began as a biennial effort—first in 1983 and then again in 1985. In 1987, the magazine embarked on a far more ambitious undertaking. Once again it only surveyed presidents, but this time it published the results not just in the magazine but also in a separate guidebook called "America's Best Colleges." That first guidebook included rankings of law, business, medical, and engineering schools.

[2]As such, some organizations or individuals are not individually named.

The college guidebook is published in late August or September, at the same time that the rankings appear in a regular issue of the magazine. Starting in 1997, *U.S. News* put online substantially more information than there was room for in the printed guidebook. It now charges a fee to see the full rankings, although much of the information is available without any cost. The timing of the publication of the college rankings is based on the belief that they should come out at about the same time that high school students are starting back to school and, presumably, thinking about college.

After including graduate school rankings in the initial guidebook, the magazine published them only in a regular issue of the magazine for several years. It was unwilling to produce a separate graduate school guidebook until it was able to obtain a lead advertiser to underwrite the costs. Despite the success of the college venture, the management of the magazine took a very conservative stance when it came to expanding what quickly developed into a franchise product for the company. In 1994, *U.S. News* finally found a lead advertiser for a graduate school guidebook, the Chrysler Corporation. It has published an annual graduate school guidebook ever since, although the advertisers have changed over the years. Similarly, the advertisers for the college guidebook have changed, although for a long time State Farm Insurance was the lead advertiser.

The graduate school rankings come out in March or April, about six months after the college rankings. It is impractical to try to do both the college and graduate school rankings at the same time—there is simply not enough staff to handle both projects simultaneously.

U.S. News's first guidebook, while published in 1987, says "1988 edition" on the cover. Guidebook covers are always dated a year ahead. The reason: It keeps the guidebook, which contains far more data and editorial content than there is room for in the magazine, on the newsstands for almost a full year. In fact, in order to keep the guidebook in circulation for that length of time the magazine has to change the cover after about six months, although the content remains unchanged. The change in covers is related to the arcane rules of magazine distribution. As a result of the switch in covers, some people buy what they think is a new edition of the guidebook only to discover that the contents have not changed. In some cases they ask for, and receive, a refund.

The methodology used in the first annual rankings issue and guidebook was very simple. At the undergraduate level, presidents were asked to pick the 10 schools in their academic category that did the best job of providing an undergraduate education. To reflect the diversity of American higher education, institutions were placed in one of nine categories: National Universities; National Liberal Arts Colleges; Smaller Comprehensive Institutions; Southern Comprehensive Institutions; Eastern Comprehensive Institutions; Western Comprehensive Institutions; Western Liberal Arts Colleges; Southern Liberal Arts Colleges; and Eastern Liberal Arts Colleges. The academic categories were based loosely on classifications established by the Carnegie Foundation for the Advancement of Teaching, whose categorization of higher education institutions is a staple of academic research.

The magazine published a ranking of the top 25 institutions in the National University and National Liberals Arts College categories and the top 10 in the other categories.

Action and Reaction

With the publication of that first guidebook, whose dimensions are roughly similar to those of a hardcover dictionary, college presidents and other administrators began to realize that the

rankings were no longer just an occasional survey that would appear only in the weekly magazine. It was taking more permanent form, something people could put on their bookshelves.

As leaders of American higher education began to take notice, many of them did not like what they saw. They viewed the rankings as nothing more than a beauty contest and urged the magazine to halt the project. They felt the magazine's approach was not a suitable way to assess America's complex and multifaceted system of higher education.
A number of presidents asked for a meeting with the magazine's editors to express their concerns. That meeting led the editors to conclude that if the rankings were to be credible and have staying power, a major transformation was needed. After consulting with a variety of experts, the editors made two major changes in the rankings. First, the universe of those surveyed was expanded to add college and university chief academic officers/provosts and deans of admission. Second, a multidimensional methodology that made substantial use of objective data was developed.

College presidents and other administrators began to realize that the rankings were no longer just an occasional survey that would appear only in the weekly magazine.

In the 1989 edition of the guidebook, the changes in methodology were explained this way: "First, because academic deans and admissions officers often see education from rather different perspectives than do college presidents, they also have been included in the survey of more than 1,000 college officials. Second, because the expert opinions are just that, opinions, *U.S. News* has based its latest academic rankings on objective data as well as on the subjective judgments in the survey."

The objective data were initially divided into four broad categories: student selectivity, faculty quality, institutional resources, and student retention. Over the years, other categories have been added, the weights used for different components have been changed, and the titles of some categories have been altered.[3]

Within each of the four broad objective categories there were subcategories. For example, student selectivity included acceptance rates, the average standardized test scores of a college's entering class on either the SAT or ACT exams, and high school class-rank data. Each of these components had a specific weight that when totaled equaled the weight assigned to the student selectivity category.

In the 1989 edition, in addition to providing an overall ranking of schools, the magazine published a separate reputation ranking of National Universities and National Liberal Arts Colleges. There was considerable variation between the overall ranking and the reputation survey—public institutions fared much better in the reputation survey. Ten public universities were in the top 25 in reputation, but only five made it to the top 25 in the overall rankings. The relative underperformance of public institutions in the institutional resources and student retention categories helps to explain that. Even today, the objective data work to the disadvantage of public institutions. In the most recent rankings, only four public universities were in the top 25 among National Universities. But if the magazine were to do a ranking based purely on academic reputation, at least five more publics would be in the top 25.

[3]Much of this will be addressed later in the chapter.

In addition to the separate reputation ranking, the magazine used the four clusters of objective data to develop lists of the top five schools for each institutional category. For example, among national universities, Brown University ranked first in selectivity, the California Institute of Technology was first in faculty quality as well as in financial resources, and Duke University ranked first in retention. By taking this approach, the magazine made the best of a difficult situation. Because of constraints imposed by the College Board, the organization that provided the objective data for that edition, *U.S. News* was unable to publish the actual data for each school.

Change is a Constant

The changes made in that second annual edition were just the beginning of an evolutionary process that has led to continual reshaping of the methodology and many other aspects of the rankings. The rankings remain a work in progress, although an effort has been made to stabilize the methodology.

The changes made in that second annual edition were just the beginning of an evolutionary process that has led to continual reshaping of the methodology and many other aspects of the rankings.

Starting with the 1990 edition, the magazine obtained data from an organization that did not place constraints on publication of the objective data. As a result, the magazine was able to show where schools ranked both overall and in reputation, as well as in the categories based on objective data. All that information was incorporated into one ranking table for each category of institution. The table for National Universities showed that Yale University ranked first overall, and fifth in academic reputation, third in student selectivity, third in retention, second in faculty quality, and ninth in financial resources.

Even though fewer than 200 institutions were actually ranked, the 1990 edition introduced a comprehensive table that contained data for over a thousand institutions, providing students and families with information they could use to compare colleges and universities that were not ranked. The editors felt that since the information was being collected for the rankings, why not provide it to the public.

That year's edition also introduced a new and controversial element. Instead of just ranking the top schools in each category, each ranked school's actual score was published. Under the magazine's system, the top school in each institutional category received a score of 100. The score of all the other institutions in that category were based on how they fared in relation to the top school. Some schools were separated by only one-tenth of a point. Duke University, which ranked fifth, had an overall score of 94.3, while Stanford University, which ranked sixth, had an overall score of 94.2. It was a distinction without any real difference, and it raised the hackles of many higher education administrators already hostile to the rankings concept.

Yet, the practice of ranking schools down to a tenth of a point continued for years. Even within the magazine there was ongoing debate about the practice. Outside critics argued that this practice created an illusion of false precision, while defenders within the magazine felt that eliminating the decimal point would lead to more ties and thus would be unhelpful to the rankings. There were those within the magazine who did not embrace that view, but it was not until the 1998 edition

that the practice was changed. Ever since then, scores have been rounded off to the nearest whole number. This did create more ties. For example, in the 1998 edition five schools tied for ninth place in the National University rankings, and there were a number of other ties among the top 25. But whatever might have been lost by no longer ranking schools down to one-tenth of a point was more than offset by the credibility and goodwill generated by making the widely desired change.

Opening the Black Box

Developing the methodology for the rankings was really a process of trial and error. In the 1991 edition, the magazine began to shed light on the weighting scheme used in its rankings. At that point, the magazine was still using four broad categories of objective data plus the subjective reputation component. Over the years the number of objective data categories has expanded to seven.

In the 1991 edition, the magazine explained its weighting scheme this way: "Because most experts believe an institution's student body, faculty, and reputation are the major components of what makes a 'best college,' these…were each given weights of 25 percent in determining overall rank. Financial resources counted for 20 percent and student satisfaction for 5 percent." Student satisfaction was actually a misnomer. It really was the five-year graduation rate average for three preceding classes. That "output" measure has since been renamed "retention" and now includes both a school's six-year graduation rate and its freshman retention rate. It has come to be weighted more heavily over the years for reasons that will be discussed elsewhere in this chapter.

The explanation of the methodology in the 1991 edition was the beginning of an ongoing effort to spell out clearly how the rankings are determined. The aim was—and is—to make the methodology transparent. Experts could quarrel with the methodology—and quarrel with it they did—but they could not accuse *U.S. News* of operating behind a curtain of secrecy. Editors believed that for the rankings to be credible, it was essential to make the methodology transparent.

The explanation of the methodology contained this cautionary note: "As in previous years, the complex methodology was developed after consultations with scores of college presidents and other academic officials. Because of changes made as a result of their recommendations, the methodology continues to evolve and, therefore, the 1991 rankings are not directly comparable to those published in previous years."

That warning remained valid more often than not as the magazine continually tinkered with the weights given to the objective data and introduced new data categories. Throughout this process, one constant has been the weight given to the subjective component of the methodology, institutional reputation, which is now called "peer assessment." It accounts for 25 percent of a school's ranking.

While those within the magazine have always known that year-to-year comparisons of a school's ranking are not valid in years when changes are made in the methodology, it has not always stressed that point in talking with other news organizations. And even if it had done so, it is almost inevitable that press and higher education institutions would still have made comparisons. Moreover, colleges and universities that did well in the rankings began to tout them, without regard to the fact that the methodology might have changed.

While the magazine was frequently accused of changing the methodology to shake up the rankings, nothing could be further from the truth. The methodology was changed in response to valid criticism from outside experts.

Almost from the beginning, editors met regularly with college officials. As a consequence, it was not unusual for the editors to meet with as many as three presidents in a single day. Frequently, presidents complained that the ranking formula put too much weight on "input" variables such as student selectivity and not enough weight on "outcome" variables such as freshmen retention and graduation rates. Their basic argument was that input variables show how smart incoming freshmen are, but do not take into account how good an educational job colleges are doing with the students who enroll. As a result of the feedback, in the 1996 edition the magazine made a significant change in its methodology.

Lacking any comprehensive comparable data on what students actually learn during their undergraduate years, the editors decided that freshmen retention rates and graduation rates were the best proxy available for evaluating outcomes. They decided to reduce the weight placed on student selectivity from 25 to 15 percent of a school's overall ranking. They shifted the remaining 10 percent to retention, which includes the freshman retention rate and an institution's six-year graduation rate. In short, they put more emphasis on outcomes than inputs. That created some changes in the rankings, but it did not precipitate an upheaval that could have undermined the credibility of the project. In fact, the editors pretested the change in weights to make sure that it would not produce an upheaval.

This was not the first effort to put greater weight on outcomes. In the 1994 edition, the magazine had added a category called alumni satisfaction, the percentage of living alumni who gave money to their institution's fund drives in two preceding years. The editors viewed this measure, which accounted for five percent of an institution's overall ranking, as a very rough proxy for how satisfied graduates were with the education they received. They saw this as an outcome measure, although critics felt it was more a measure of the effectiveness of an institution's development office, a criticism not without validity. But inclusion of the data caused some institutions to keep better track of their alumni and to step up their fundraising.

Another effort to focus on outcomes was the development of a concept called "value added." That measure, introduced in the 1997 edition, accounted for 5 percent of an institution's ranking. It was limited to the National University and National Liberal Arts College categories. The 5 percent was subtracted from retention, which was reduced from a weight of 20 to 25 percent.

The new measure was designed to show the effect of a college's programs and policies on the graduation rate of students, after controlling for spending and student aptitude. The measure was rooted in the idea that schools that admit the brightest students, as measured by SAT or ACT scores, and, to a lesser degree, that expend the greatest amount of money per student, should have higher graduation rates than peer institutions that admit students with lower scores and have less money to spend on their education. But the data show that some schools with less academically gifted students and lesser resources do as good a job of retaining and graduating students as more well-endowed institutions that enroll students with higher test scores. Again, this is not a perfect measure, but it represents an effort to address complaints from a number of institutions with lesser resources that the rankings unfairly rewarded well-endowed institutions that enrolled students with higher test scores. The basic concept of the value-added measure has remained

intact, although the statistical approach has changed somewhat. Today, the measure is called "graduation rate performance" rather than value added.

Another change occurred in the 2004 edition, when *U.S. News* stopped using "yield"—the proportion of admitted students who enroll in a college—in its ranking formula. Yield counted for only 1.5 percent of an institution's overall ranking, but the attention it received left the impression that it was far more heavily weighted. Many critics of the rankings claimed that by including yield in its methodology, *U.S. News* was contributing to what had become an explosion in early-decision applications at many selective institutions. They felt that early decision, under which students apply to one school early in their senior year and are committed to attend if accepted, was skewing the way students went about applying to college.

Critics argued that the magazine was encouraging colleges to expand early decision since the yield to a college from students admitted early is guaranteed. If a college were to admit almost half its class early, as some do, that would strengthen its yield and thus conceivably help its ranking. Critics felt that while that might be good for colleges, it was not good for many students, who were making college decisions prematurely.

Tired of being blamed for the expansion of early decision, *U.S. News* decided to drop yield from its equation. That change had minimal impact on the rankings since yield counted for so little to begin with. It also had little impact on early decision, which shows no signs of diminished popularity. But it did take the magazine out of the crosshairs of those who dislike early decision and who sought to make the magazine the villain for a development that reflected the increasing intensity of college admissions.

Expanding the Rankings

In the first several years of the annual rankings, the magazine limited the number of schools ranked. But as confidence in the rankings increased, the magazine began to rank more and more schools. Beginning with the 1992 edition, it placed schools that ranked below the top group in quartiles within their institutional category and listed the schools in each quartile alphabetically. That provided a relative sense of where all schools ranked in their category. Starting with the 1996 edition—by then the magazine was doing its own data collection instead of relying on outside contractors—*U.S. News* began to number-rank more schools. For example, in the National University category that year it ranked the top 50 institutions, instead of just the top 25.

As time passed, the number of schools that were ranked continued to expand. Today, the magazine ranks the top 50 percent of colleges and universities in each category of institution. The rest of the schools in that category are grouped in two tiers and listed alphabetically. In the National University category, that means about 125 institutions receive a numerical ranking; about 110 are number-ranked in the National Liberal Arts College category, and so on.

In theory, the magazine could number-rank every institution. But years ago the editors made a decision not to do that. There were two major reasons for this. First, as a general rule, the further down a school is in the academic food chain, the less likely it is to possess all the objective data used in the rankings methodology, or, if it has the data, it is less likely to be willing to provide the information. Consequently, a precise ranking of schools lower down in the pecking order is harder to achieve.

When a school fails to provide data, *U.S. News* tries to obtain the information from other sources. If it cannot do so, it has utilized a variety of statistical methods to estimate the data. The magazine indicates the source of the data when it does not come from the institution itself and explains the statistical method used to make estimates. But the very fact that it makes such estimates is not without controversy.

A precise ranking of schools lower down in the pecking order is harder to achieve.

Perhaps the most heated controversy arose in 1995, when Reed College in Portland, Ore., a highly regarded liberal arts institution, chose to boycott the rankings by not submitting any data to the magazine. In the first year of the boycott, Reed was given the equivalent of the lowest score among National Liberal Arts Colleges for each piece of objective data used in the magazine's methodology. As a result, Reed ended up in the bottom tier, an outcome that was more punitive than logical. The next year *U.S. News* collected as much data from outside sources as it could and used a different estimating approach for data that were missing. As a result, Reed moved into the top 40 among National Liberal Arts Colleges. Reed's boycott continues today, as does the magazine's effort to collect data on Reed from other sources. That effort is made easier by the fact that Reed publishes institutional data collected by what is known as "the common data set" on its own Web site.[4]

The second reason for not number-ranking all institutions is that the editors felt that if the magazine were to do so, a few institutions would inevitably end up being labeled as the worst colleges in America. *U.S. News* had no interest in that outcome, which could have detracted from the overall rankings. The goal of the project from its inception has been to identify the best schools, not to single out the worst.

Under Fire

As the rankings became more visible and were perceived to have significant impact on decisions by students and their parents on where to apply to college and even where to enroll, they came under more fire. Academic experts complained about the weighting scheme used in the rankings. They argued that despite changes made in the methodology it remained arbitrary and lacked any empirical basis. Consultants from NORC, a research firm hired by the magazine to assess the rankings methodology, agreed with the critics. They concluded in a 1997 report that "the principal weakness of the current approach is that the weight used to combine the various measures into an overall rating lacks any defensible empirical or theoretical basis." The researchers went on to say that criticism of the weighting did not mean that the weights used were necessarily wrong, but that "it is difficult to defend on any grounds other than the *U.S. News* staff's best judgment on how to combine the measures." The debate over the appropriate weights is no closer to resolution now than it was then.

Other criticism came from academic leaders who said that it was impossible to measure the quality of institutions. They felt that each institution was distinctive and the magazine's attempt to reduce a college to a set of numbers was an exercise in measuring the unmeasurable.

[4]The common data set was established in the mid-1990s to standardize data submitted to *U.S. News* and other guidebook publishers and is run by the participating publishers.

High school counselors joined in the attack. They agreed with the critique that the magazine was trying to quantify the unquantifiable. But their core criticism was that the rankings, viewed by *U.S. News* as a good faith effort to help students and families sort through the complex process of choosing a college, actually did a disservice to students and parents. They argued that the rankings were causing students to look only at a school's status, while ignoring what counselors call "the fit" between a prospective student and an institution. They felt, for example, that students who were unconventional in their outlook were not necessarily considering schools where they would be most comfortable socially. Instead, they were concentrating on schools that stood high in the rankings.

The argument that most families mindlessly followed the rankings without regard to other considerations was—and remains—debatable, and the editors were willing to engage in the debate. They felt it was important to present their point of view and to obtain constant feedback, even if the experience was often painful. They routinely accepted invitations to speak before organizations of educators, where they usually got an earful. One college admission official at a national meeting went so far as to compare *U.S. News* to Satan.

As painful as these confrontations were, they had a salutary effect. In an effort to improve relations with critics and to make the rankings more useful to students and parents, *U.S. News* established two advisory committees, one of college admissions deans and another of high school counselors. It has since added a third committee composed of institutional researchers. Editors meet with the committees annually to discuss the rankings and issues related to admissions and financial aid. These meetings are not an exercise in public relations. They involve substantive discussions that yield concrete results.

For example, starting with the 1995 edition the editors emphasized that "rankings are just one of many criteria prospective students should take into account in choosing a college. Simply because a school ranks at the top of its category does not mean that it is the best choice for every individual. A student's academic and professional ambitions, financial resources, and scholastic records, as well as a school's size, atmosphere, and location should play major roles in college choices."

In the 1997 edition, the magazine went a step further. Editors worked with counselors to develop profiles of hypothetical students with different academic credentials, outside interests, and financial and personal situations. They then asked counselors around the country to come up with a list of colleges that would be suitable for each individual. Many of the suggested schools were not highly ranked. The idea was to illustrate in a concrete way that there are factors other than rankings to consider when choosing a college.

Researchers have found that the rankings play a role in whether students apply to and enroll at a college or university, but there is disagreement about the magnitude of the impact. A paper by researchers at UCLA concluded that "academic reputation is a powerful influence on students, more powerful than the advice of professional advisors or the influence of families (McDonough, P., et al. 1997). We believe that colleges and newsmagazines need to take actions to place rankings in a broader decision-making context for the students, parents, and professional advisors who are relying on them."

Researchers James Monks and Ronald Ehrenberg looked at the experience of 31 selective private colleges and universities and found that when an institution improved in the rankings, the following year it received more applications, accepted fewer applicants, and enrolled a higher

proportion of those accepted. Conversely, if it fell in the rankings, the reverse occurred. Their findings were published in the November/December 1999 issue of *Change* magazine.

The 2005 Survey of College Freshmen by the Higher Education Research Institute at UCLA provided a different perspective. Of the reasons freshmen cited as very important in influencing their decision to enroll at a school, "rankings in national magazines" was tenth among 18 possible factors. It was cited by just 16.6 percent of the more than 263,000 students around the nation who participated in the survey.

More recently, a study of 600 high-achieving high school seniors who entered college last fall found that rankings stood near the bottom of the list of information sources and of more general factors that play a role in whether a student applies to a college (Lipman Hearne, 2006). However, the study, done by a Chicago-based market research firm that has many higher education clients, did find that students in several Mid-Atlantic states—New York, Pennsylvania and New Jersey—were much more likely to take the rankings into account than students in other regions.

Whether the impact of the rankings is as great as some critics believe or as modest as some of the research suggests, the bottom line is that the rankings have real-world impact.

Grappling with Crises

The initial outcry over the rankings in 1987 was the first in a series of crises that have occurred over almost two decades. Some were relatively minor and of little interest to those outside the magazine, but a few have been significant.

Arguably, the most important occurred in 1995 after *The Wall Street Journal* published a front page story in its April 5 edition headlined "Colleges Inflate SATs and Graduation Rates in Popular Guidebooks" (Stecklow). The story was more about the games colleges played with their data than about the *U.S. News* rankings per se, but the well-documented report raised basic questions about the accuracy of some of the objective data on which the rankings were largely based.

The author of the story, Steve Stecklow, compared the data that colleges reported to debt-rating agencies, who rate their bond issues, with data they submitted to *U.S. News* and other publishers. He found that there were significant discrepancies in SAT scores, acceptance rates, and graduation rates. The information sent to the magazine was almost always more favorable to the school than the data sent to the debt-rating agencies, Standard & Poor's, and Moody's Investors Service, Inc. Stecklow pointed out that lying to the magazine carried no penalty, but lying to the debt-rating agencies could lead to federal penalties.

His story showed that some colleges excluded the scores of certain groups of students to improve their SAT average. For example, Northeastern University in Boston had excluded both international students and remedial students. That boosted the school's average SAT score by 50 points. A number of other schools cited in the story employed similar tactics; some claimed it was not purposeful, but just a mistake.

What the schools were doing was ethically indefensible. From the magazine's standpoint, it was potentially damaging to the credibility of the rankings. Realizing that the future of the rankings might be at stake, *U.S. News* revised the questionnaire sent to colleges and tightened definitions. It

attempted to insure that schools did not omit test score data for any group of students by asking for data for all first-time full-time freshmen, including those in special programs. A new question specifically asked schools whether they had complied with the instructions for providing test score data. If they acknowledged that they were omitting scores for some students, *U.S. News* then did its own statistical estimate for ranking purposes. In another effort to try to assure accuracy, the magazine programmed its computers to flag significant discrepancies in a school's data from one year to the next. If it found a discrepancy, it contacted the school seeking an explanation.

Realizing that the future of the rankings might be at stake, *U.S. News* revised the questionnaire sent to colleges and tightened definitions.

The magazine also began to cross-check data available from other sources, including Moody's, the debt-rating agency; the American Association of University Professors, which collects data on faculty salaries; the National Collegiate Athletic Association, the organization that oversees college sports; and, more recently, the U.S. Department of Education. For example, it checked graduation rate data submitted by colleges to comparable data collected by the NCAA. Institutions that submit misleading data to the NCAA face potential penalties. When the magazine found discrepancies in data, it contacted the school involved. If it could not get a credible explanation, it used the data submitted to the NCAA and said in a footnote that it was doing so. Today, it no longer contacts schools when it finds discrepancies. It simply uses the data submitted to outside organizations and explains that in a footnote.

In the 1996 edition of the guidebook, *U.S. News* laid out the steps it had taken to ensure the integrity of the data. One college president suggested to the editors that they ask schools to have their auditors attest to the accuracy of the data submitted to the magazine. The editors liked the idea, but felt that the magazine was in no position to make such a demand.

Even with the steps that *U.S. News* has taken and with increased requirements at the federal level for institutions to provide standardized data, it is possible that some schools might still be playing games with data. Some institutions use a two-part application and might count those students who fill out only the first part as applicants even though that is contrary to the instructions in the magazine's questionnaire. By counting only those who fill out the first part of the application, a school increases the size of its applicant pool and, as a result, appears more selective than it is. For example, if a school counts 20,000 students who completed part one as applicants, and admits 10,000 of them, its acceptance rate is 50 percent. But if only 15,000 students completed both parts one and two, then the school would have fewer actionable applicants and its acceptance rate would rise to 67 percent. The difference between a 50 percent and a 67 percent acceptance rate could impact an institution's ranking.

There are other games schools can play, including admitting students with weaker academic profiles for the second semester, thus not having to include them in the data reported to the magazine, which asks for information for first-time full-time freshman who entered in the fall semester.

U.S. News can do the most careful job possible, but institutions that are determined to present misleading data can find a way to get around tightened definitions and other steps that the magazine has taken. The fact that some institutions are willing to engage in such maneuvers, the equivalent of trying to find a loophole in the tax code, says a great deal about the perceived stakes.

A second crisis occurred in 1999 when a new director of data research insisted on abolishing a statistical procedure that, as the magazine put it, had "flattened out" large disparities in one particular data point used to rank National Universities. The data point in question was expenditures per student. The flattening out procedure had been used because the California Institute of Technology, a small institution with a large research budget, consistently reported expenditures per student that were more than double those of the school that ranked immediately behind it in expenditures. The gap was so large that it would have skewed the rankings had an adjustment not been made. Once the adjustment was dropped, Caltech vaulted to the top of the 2000 rankings; it had ranked ninth the year before. The outcome was predictable for anyone familiar with the magnitude of the disparity.

There are other games schools can play, including admitting students with weaker academic profiles for the second semester.

But the results came as a surprise both to some within the magazine and to the outside world. Editors had to do a lot of explaining to justify what seemed an implausible change in the rankings. The next year, after the director of data research in question had departed the magazine, *U.S. News* instituted a new statistical adjustment that dampened the impact of the vast disparity in per student expenditures. As a result, in the 2001 rankings, Caltech dropped to fourth, and in the most recent rankings it was seventh.

Both of these events had a common thread: they called into question the credibility of the rankings. The magazine managed to weather both crises, but it was not easy.

Why the Success of the Rankings

The phrase "timing is everything" is a cliché, but that makes it no less true. It certainly applies to the *U.S. News* rankings. They came along as the consumer movement in America was reaching full flower. A generation of parents who were college-educated brought both pragmatism and status-seeking to the college search process.

While many members of earlier generations were simply pleased that their children were going to college, members of the Baby Boom generation cast a more critical eye toward higher education. They wanted value for their money. Some viewed higher education as a commodity and wanted the biggest possible bang for their buck, especially as the cost of going to college increased far more rapidly than family income. Others felt it was critical to go to "the best" college, by which they meant the most prestigious, an institution whose decal they could proudly put on the rear window of their cars or SUVs.

For these families, the rankings have provided a source of comparable information to help them navigate the complex task of identifying suitable colleges for their children. The rankings have become especially important for those with children in public schools at which college counseling is inadequate.

But it is not consumers alone who have helped fuel the success of the rankings. Colleges and universities have played a major role. Even as they have decried the rankings, many institutions have used them to market themselves, especially when they have done well, and sometimes even when they have not. One institution took out an ad in an airline magazine touting itself as being ranked among America's best colleges by *U.S. News*. That was literally true. The institution was

ranked in *U.S. News's* "America's Best Colleges." But the ad was totally misleading since the school in question ranked in the bottom tier of its particular category. However, the fact that even a bottom tier school sought to use the rankings to market itself illustrates how higher education has increased the visibility and impact of the rankings. In fact, *U.S. News* itself did very little marketing. Why spend money when others do the marketing for you?

Colleges and universities also have used the rankings for internal benchmarking and to see how they compare to other institutions with which they compete for students, research dollars, and status. Sometimes the impetus has come from presidents, who turn to the rankings to see whether they have met the goals they have set for their administration. Other times the impetus has come from members of an institution's board of trustees. Often they are members of the business community where, unlike higher education, success is defined in concrete, quantifiable terms. For these board members, the rankings offer the kind of assessment that they are accustomed to and rely on.

The Future

As increases in tuition have continued to outstrip growth in family income, there has been stepped up political pressure on colleges and universities to provide comparable data to help families evaluate the institutions. There has even been discussion by members of a federally appointed higher education commission of requiring a standardized test at the beginning and end of students' college careers to see just how much they have learned, with the results made public on an institution-by-institution basis.

Much of higher education is opposed to what administrators and faculty members view as intrusive steps. They believe that America's colleges and universities are just too diverse to use a standard measure or set of measures to evaluate institutions. But pressure on institutions to come up with one or more ways to demonstrate the value of the education they provide and to do so on a standardized basis is not likely to diminish, regardless of which political party is in power.

Until and unless higher education institutions can come up with their own "rankings," which need not be a literal numerical ranking in the *U.S. News* tradition, but rather an easily understandable and quantifiable way for consumers to compare institutions, *U.S. News's* rankings are likely to continue to flourish. To do so, however, they must remain credible and the magazine's editors must remain open to modifying the methodology if better ways to assess institutional quality and value emerge.

References

Lipman Hearne. 2006. High Achieving Seniors and the College Decision, April 2006.

McDonough, P., et al. 1997. College rankings: Who Uses Them and with What Impact. Annual Meeting. American Educational Research Association.

Monks, J. & Ehrenberg, R. G. 1999. *U.S. News & World Report's* College Rankings: Why do They Matter? *Change*, Nov/Dec 1999, 31(6), 42-51.

Stecklow. Steve. 1995, April 5. Cheat Sheets: Colleges Inflate SATs and Graduation Rates in Popular Guidebooks. *The Wall Street Journal,* A1, 8.

A Global Survey of Rankings and League Tables

By Alex Usher and Massimo Savino[5]

University rankings or "league tables," a novelty as recently as 15 years ago, are today a standard feature in most countries with large higher education systems. They were originally created over 20 years ago by *U.S. News & World Report* in order to meet a perceived market need for more transparent, comparative data about educational institutions. Reviled by critics but popular with parents, copycat ranking systems began popping up all over the world, usually shortly after the introduction of—or a rapid rise in—tuition fees. Wherever rankings have appeared, they have been met with a mixture of public enthusiasm and institutional unease.

One of the main causes of institutional unease is the tendency of institutional ranking schemes to use weighted aggregates of indicators to arrive at a single, all-encompassing quality "score," which in turn permits institutions to be ranked against one another. By selecting a particular set of indicators and assigning each a given weight, the authors of these rankings impose a specific definition of quality on the institutions being ranked. The fact that there may be other legitimate indicators or combinations of indicators is usually passed over in silence. To the reader, the author's judgment is in effect final.

Intriguingly, however, there is little agreement among the authors of these indicators as to what indicates quality. The world's main ranking systems bear little if any relationship to one another; they use very different indicators and weightings to arrive at a measure of quality.

This chapter examines 17 university league tables and ranking systems from around the world. Fourteen of these are "national" league tables collected from nine countries (Australia, Canada, China, Hong Kong, Italy, Poland, Spain, the United Kingdom and the United States); three are "international" or "cross-national" league tables. Specifically, these league tables are compared in terms of their methods of data collection and their selection and weighting of indicators. We also look at an eighteenth ranking system (the German Centre for Higher Education Development rankings), which does not conform to the standard league table "rules."

What Are University Rankings and League Tables?

University rankings are lists of certain groupings of institutions (usually, but not always, within a single national jurisdiction), comparatively ranked according to a common set of indicators in descending order. University rankings are usually presented in the format of a "league table," much

[5]This chapter is condensed and enhanced from a larger report authored by Usher and Savino, *A World of Difference: A Global Survey of University League Tables,* Educational Policy Institute: January, 2006.

as sports teams in a single league are listed from best to worst according to the number of wins and losses they have achieved.[6]

Another notable aspect of league tables is that they are, for the most part, produced by commercial publishing enterprises. In part, this reflects the fact that rankings share some characteristics with "consumer guides" to various products. Although rankings are not guides to specific institutions, the publishers of individual institutional guides may incorporate rankings data as supplementary material, fleshing out descriptions for the purpose of providing more information to their readers. Rankings are—at least in theory—meant to be an "under the hood" look at a complex product. In many cases, the effort required to collect, collate, and analyze the data required to produce the rankings is so great that their production on anything but a commercial basis is probably impossible.

University ranking systems come in two varieties: institutional ranking systems and sub-institutional ranking systems. They can be conducted either on a national or international scale. National ranking systems are ones in which all or nearly all of a country's universities are measured against one another. This was the original university ranking format—i.e., the type pioneered by *U.S. News* in 1981 and which has been widely copied in other countries. In most cases, all universities within a country are compared, although in some cases—notably in Canada (*Maclean's Magazine*) and the United States (*U.S. News*)—the country's universities are divided up according to certain institutional characteristics and only compared to other institutions with similar characteristics, in effect creating a group of mini-league tables.

Global institutional ranking systems are a new variation on the older idea of national rankings. There are at present only two of these: the *Academic Ranking of World Universities* from Shanghai's Jiao Tong University, first released in 2003, and the *World University Rankings* from the *Times Higher Education Supplement* of Britain (henceforth *THES*), first released in November 2004. The first international ranking—albeit not a global one—was actually done by *Asiaweek* magazine in 1997, which ranked the continent's major universities.

Beyond institutional rankings, there are also sub-institutional rankings, which compare specific university units against similar ones at other institutions. These rankings are usually national in scope and deal with professional schools such as business, law, and medicine. Graduate business schools are also the subject of a number of international rankings, from such organizations as the *Economist,* the *Financial Times,* the *Wall Street Journal* and *Business Week.* These types of league tables are not covered in this chapter, on the grounds that there are simply too many of them to analyze in detail. However, one variation on the subject-specific ranking system (the CHE rankings) will be examined at the conclusion of the chapter.

There are also ranking schemes that focus on specific aspects of university activities. For instance, the *Best American Research Universities* ranks U.S. institutions specifically on their research output, as, in a cruder manner, does the Centre for Science and Technology Studies in Bern, Switzerland, with its international "Champions League" tables. Similarly, *Yahoo Magazine* has ranked universities on their "connectivity," and *Diverse: Issues in Higher Education* (formerly known as the *Journal of Blacks in Higher Education)* has graded them on their ability to integrate students from

[6]The term stems from U.K.-based chart listings that were often compared with Premier League professional soccer or football standings in England during the 1990s and can now be found in an extremely wide variety of contexts in Britain today.

different backgrounds in its ethnic diversity rankings. These types of ranking systems are excluded because their purposes are much more specific and limited than the general ranking systems that are the focus of this review.

How Rankings and League Tables Work

League tables, by their very nature, are meant to boil down the work of entire institutions into single, comparable, numerical indicators. In most ranking systems, this comparison is a three-stage process: first, data are collected on indicators; second, the data for each indicator are scored; and, third, the scores from each indicator are weighted and aggregated.

All rankings systems operate by comparing institutions on a range of indicators. The number of indicators in a ranking system can vary significantly, from five in the simplest case (the *THES World Rankings*) to several dozen in the case of the most complicated (*La Repubblica or Wuhan*). Specific areas of institutional activity or types of institutional output can therefore be compared across institutions, in much the same manner as is done with performance indicators.

League tables, by their very nature, are meant to boil down the work of entire institutions into single, comparable, numerical indicators.

With only a few exceptions (notably, Spain's *Excelencia* rankings), league table systems then take the data on each indicator and turn them into a "score." Usually, this is done by giving the institution with the highest score on a particular indicator a perfect mark of 100 and then awarding lower scores to other institutions based on how close they were to the score of the top institution. Once scores have been derived for each indicator, they are weighted, with greater weight being accorded to indicators that are believed to be of greater importance. The weighted scores from all indicators are then tallied to give a unified final score for each institution.

Clearly, the choice of indicators and the weight given to each indicator makes an enormous amount of difference in the final output. Indeed, it is no exaggeration to say that when publishers advertise their product as a guide to "the best" institutions, it is the publishers themselves who largely decide the best simply through their choice of indicators and weightings. In effect, the act of choosing a set of indicators and weightings imposes a one-size-fits-all definition of "quality."

The Evidentiary Basis of League Tables—How Data Are Collected

A key issue in the preparation of league tables and rankings is the method by which data are collected. There are basically three sources of data on institutions:

- *Survey data.* Surveys of the opinions or experiences of various stakeholders can be used to obtain comparable data on different institutions regarding educational quality.
- *Independent third parties.* Frequently, government agencies will collect and publish data on institutions in their jurisdiction that can be used as an objective standard by which to compare institutions. These data are often financial in nature and based on administrative data from grant-making bodies.
- *University sources.* The most complete and most detailed sources of data on universities are of course universities themselves, and they are thus potentially a very rich source of data.

The use of each source of data has pros and cons. Survey data are scientific in the sense that it records observations accurately, but to the extent that it is used to survey employers or opinion-makers on the value of degrees from various institutions, critics might reasonably question the value of such observations, as few employers or opinion-makers are likely to have detailed views on or knowledge of every institution under scrutiny. Surveys of students and recent graduates are similarly denigrated on the grounds that while they may be able to enunciate their feelings about their own institution, they have no basis on which to compare their institution with others.

Independent third-party administrative data (usually from governments or grant-making bodies) are generally considered the "gold standard" of comparative data since they are, at least theoretically, both accurate and impartial. The problem is that these data are not (usually) collected for the purpose of compiling league tables but rather as an administrative by-product of ordinary business. As a result, over-reliance on this source of data can lead to a situation where indicators are chosen simply on the basis that data are available rather than because they contribute to a sensible definition of quality.

Finally, there are data from universities themselves. In some cases, where important indicators of quality cannot be obtained via surveys or third parties, the authors of ranking schemes will address a questionnaire to institutions and ask for certain pieces of data. The benefit of this approach is that one can—in theory—answer a number of questions about quality that cannot otherwise be answered. The main drawback is that there is absolutely no guarantee that institutions will actually report the data to the ranker on a consistent basis, as all have a clear incentive to manipulate data in a manner which will benefit them. Indeed, at some institutions in the United States, there are staff positions within institutional research offices that require the incumbent to do nothing but provide institutional data to *U.S. News* in a favorable light.

The extent to which each ranking system uses each source of data is shown in Table 1, which shows that surveys are the least frequently used source of data for indicators. Indeed, only Hong Kong's *Education18* rankings come close to having a plurality of indicators come from this source. This measure somewhat underestimates the importance of surveys, however, as it does not account for the weighting given to each indicator in each study. In the *THES World Rankings,* for instance, there is only a single survey (for "reputation"), but it accounts for 50 percent of the total ranking. Similarly, Canada's *Maclean's* rankings have only one survey-based indicator out of a total of 24, but this one indicator is worth 20 percent of the final score.

Outside North America, third-party sources are by far the most heavily used sources of data: indeed, four of the 18 ranking schemes listed here use them exclusively. Of the remaining 14, third-party sources comprise a plurality of indicators in eight rankings and university sources form a plurality in six. The predominance of data from universities is most understandable in the cases of the *Asiaweek* and *THES* rankings, as their international scope significantly reduces the possibility of third-party sources providing data on a consistent transnational basis (*Shanghai Jiao Tong,* the third international study in this comparison, solved this problem by relying almost exclusively on research output measures such as scientific publications and citations). In the cases of *U.S. News, Maclean's,* the *Guardian,* and *Rzeczpospolita,* the explanation seems to be that the editors' definitions of "quality" could not be measured using government administrative data. This may indicate a failure of government data collection in

TABLE 1: Number of Indicators by Type of Data Source

	Raw Indicator Count	Survey Data	Third Parties	Universities
Asiaweek—Asia's Best Universities (defunct, 2000)	18	–	–	18
Daily Telegraph (2003)	1	–	1	–
Education18.com	9	3	4	2
Excelencia, 2001	71	–	71	–
Financial Times (2003)	17	–	17	–
Guangdong Institute of Management Science	17	–	14	3
Guardian—University Guide 2005	7	–	2	5
La Repubblica	23	2	21	–
Maclean's University Rankings	24	1	5	18
Melbourne Institute— International Standing of Australian Universities	26	3	23	–
Netbig, 2004	18	1	10	7
Perspektywy/Rzeczpospolita Uniwersytet	18	1	2	15
Shanghai Jiao Tong University—Academic Ranking of World Universities	6	–	6	
The Times—Good University Guide 2005	9	–	9	–
Times Higher Education Supplement—World University Rankings	5	1	1	3
U.S. News & World Report—America's Best Colleges 2006	15	1	3	11
Wuhan University Centre for Science Evaluation	45	2	22	21

Source: Authors' compilation

these countries, in the sense that information deemed important to quality measurement is not collected consistently nor centrally; alternatively, it may indicate that the rankers' views of what constitutes an indicator of quality is not shared by governments or the higher education community.

What League Tables Measure—A Look at Indicators and Weightings

It should come as no surprise to learn that different ranking systems use very different indicators in order to obtain a picture of "quality." The number of individual indicators used in ranking systems worldwide runs well into the hundreds, making any kind of comparison grid too large to be useful.

In order to look at indicators and weightings in a manageable way, we have categorized them into seven larger headings, based in part on an existing model of institutional quality. Finnie and

Usher (2005), in their proposal for a system of measuring quality in post-secondary education, developed a conceptual framework for quality measurement based on the following four elements:

- *Beginning characteristics* represent the characteristics, attributes, and abilities of incoming students as they start their programs.
- *Learning inputs* come in two main types:
  - i) *resources,* both financial and material, available to students and faculty for educational ends; and
  - ii) *staff,* both in terms of the numbers but also the way in which they are deployed to teach and the learning environment they create, as measured by the amount of contact time students have with their teachers, the kinds of exams they face, etc.
- *Learning outputs* represent the "skill sets" or other attributes of graduates that culminate from their educational experiences, such as critical thinking, analytic reasoning, and technical knowledge. They also include records relating to retention and completion.
- *Final outcomes* represent the ultimate ends to which the educational system may contribute, including not only such traditional measures as employment rates and incomes but also any other outcome deemed to be important to individuals and society, such as job satisfaction, being a "good citizen," etc.

As it turns out, these four elements or categories actually encompass the majority of indicators used by the ranking systems covered by this study. However, the typology can be modified in two ways: first, by making a clearer distinction between financial resources and staff, and second by including two other sets of indicators, namely "research" and "reputation."

Rankings are more than just a collection of indicators; instead, they are a *weighted aggregation* of indicators. It is therefore important to see how they are put together and how each ranking system implicitly defines educational quality through the distribution of its weighting. Although the apparent differences between ranking systems are substantial, it turns out that there are some real and intriguing similarities among particular subsets of league tables.

Table 2 shows the differences in the indicators and weightings used by different league table systems.[7] Each row listed summarizes the distribution of indicator weightings among the seven categories of indicators described in the previous section and adds up to 100 percent. It is obvious from even the most cursory glance at this table that no two ranking systems are alike and indeed that some have virtually no areas of overlap with one another.

Despite the vastly different choices of indicators and weightings evident throughout the world, certain patterns do appear when the studies are grouped together geographically. For instance, studies from China—which has four different ranking projects—place more weight on research indicators than any other studies in the world. In the most extreme case—that of Shanghai Jiao Tong University's *Academic Ranking of World Universities*—research performance is worth 90 percent of the total ranking. This is followed by *Wuhan,* where research measures are worth 48.2 percent of the final ranking, *Netbig* (45.2 percent), and *Guangdong* (42.1 percent). Much of this

[7]In so doing, this recasts much of Nina Van Dyke's work into league tables internationally.

TABLE 2: League Table Weightings

Publication	Beginning Characteristics	Learning Inputs – Staff	Learning Inputs – Resources	Learning Outputs	Final Outcomes	Research	Reputation
Asiaweek (India/Asia)	25	28.3	10	0	0	16.7	20
Daily Telegraph (UK)	0	100	0	0	0	0	0
Education18.com (Hong Kong)	20	15	5	0	0	20	40
Excelencia (Spain)	0	25	25	25	0	25	0
Financial Times (UK)	9	19	15	10	27	20	0
Guangdong Institute (China)	0	0	0	57.1	0	42.1	0
Guardian University Guide (UK)	28	35	10	10	17	0	0
La Repubblica (Italy)	10	44.4	15.6	10	0	20	0
Maclean's (Canada)	15	20	44	5	0	0	16
Melbourne Institute (Australia)	11	3.5	11	12.6	4.8	40	17.1
Netbig (China)	12	21.8	6	0	0	45.2	15
Newsweek (United States)	10	20	10	0	0	60	0
Perspektywy/Rzeczpospolita (Poland)	8	20.5	11.5	0	0	0	50
Shanghai Jiao Tong University (Intl/China)	0	0	0	10	0	90	0
The Times Good University Guide (UK)	3.3	53.3	6.7	3.3	3.3	30	0
Times World University Rankings (UK)	5	25	0	0	0	20	50
U.S. News & World Report (United States)	15	20	15	25	0	0	25
Wuhan (China)	10.6	8.5	16.6	3.4	0.6	48.6	11.7

Source: Authors' compilation

weighting comes from counting papers and citations in bibliometric studies—studies that have a heavy bias toward the hard sciences. With the exception of *Guangdong,* which has a major focus on learning outputs (mostly graduation rates), Chinese systems also put significant emphasis on institutional reputation. In contrast, comparatively little weight is put on either resource inputs or on final outcomes. Whether this is because data on these issues are scarce or because Chinese experts genuinely consider indicators of these types to be unimportant is an open question.

Other regional patterns are also evident. Rankings of universities in the United Kingdom, for instance, completely eschew the use of reputation surveys as a means of determining quality (although *THES* places a 50 percent weighting on reputation issues). British league tables put a

much higher emphasis than league tables elsewhere on measures of staff and staff quality—on average, they put over 40 percent of their weighting in this area, as opposed to an average of just 5 percent in the rest of the world's league tables combined. The two big North American surveys—*Maclean's* rankings and those in *U.S. News*—are virtually identical in the distribution of weighting, except for the fact that the Canadian version puts more weight on resource inputs and the American version puts more weight on learning outputs (intriguingly, the general category weightings of Italy's *La Repubblica* rankings are very similar in nature to those of *Maclean's* and *U.S. News*, even though the specific indicators used are completely different).

These examples demonstrate the central premise of this paper: *different ranking systems have very different definitions of quality.* The notion of "quality" in higher education is clearly a very malleable one—some observers wish to look at outputs, while others focus on inputs. Among both inputs and outputs, there is very little agreement as to what *kinds* of inputs and outputs are important. Not only is no single indicator used across all ranking schemes, no single *category* of indicators is common, either; remarkably, none of the seven basic categories of indicators are common to all university ranking systems.

One of the only previous comparative examinations of league tables (Dill & Soo, 2004) concluded, on the basis of an examination of four sets of league tables in four countries, that international definitions of quality were converging. The findings in this chapter, which are based on a larger sample, contradict their result. Part of the reason for the contradiction lies in the fact that indicators were divided into seven categories instead of four, and therefore were always likely to find more variation. Methodological differences notwithstanding, the results still conflict.[8]

Consistency of Outcomes across League Tables

One might reasonably conclude from the foregoing analysis that measures of institutional quality are not immutable and that an institution's ranking is largely a function of what the ranking body chooses to measure. A possible example in support of this proposition is Queen's University in Kingston, Canada. In its domestic rankings (*Maclean's*), it fares very well because it attracts good students and is reasonably well-endowed and well-funded. In international rankings, it fares poorly, even compared to other Canadian universities, because its small size puts it at a disadvantage in terms of non-normalized research output measures.

Due to the plethora of ranking systems that have appeared in recent years, one can now test this proposition directly. In most countries, there are at least three separate rankings "observations" made by different national and international ranking systems (those of *THES* and *Shanghai Jiao Tong,* plus one or more domestic rankings). In those instances where one can use multiple ranking schemes to look at the relative scores of institutions in a single country, we find that certain institutions invariably rise to the top: Oxford and Cambridge in the United Kingdom; Harvard, Yale, Princeton, MIT and Stanford in the United States; Peking and Tsinghua in China; and the University of Toronto in Canada. Despite the very different weighting and aggregation schemes used by the domestic and international league tables, these institutions manage to consistently monopolize the top spots. Further down the ordinal ladder, the different rankings systems start to show greater variation (i.e., there is rarely any agreement between systems as to which university

[8]The authors of this chapter believe that had Dill and Soo looked at Asian or international ranking schemes, they too would have seen these differences and revised their conclusions.

lies in tenth position) but regardless of the ranking scheme employed, "top universities" almost always seem to come out as top universities.

This poses a serious problem for interpretation. If rankings were absolutely inconsistent across all league tables, it would be easy to dismiss the whole idea of ranking as an intellectually worthless exercise designed simply to sell newspapers or magazines. If rankings were absolutely consistent across all league tables, then we might conclude that there are probably one or two "super" indicators that are driving the overall rankings, with the remainder of the indicators essentially being "chaff" that have the effect of distracting readers and creating false differentiations. But neither of these scenarios is true—in fact, what appears to happen is that different ranking schemes provide consistent results for some institutions and inconsistent ones for others.

The simplest explanation for this is that institutional league tables don't measure what their authors think they are measuring. League tables' authors believe that each indicator is a reasonable proxy for quality and that, suitably aggregated and weighted, these indicators constitute a plausible, holistic "definition" of quality. In fact, most indicators are probably *epiphenomena* of some underlying feature that is not being measured. That is to say, there is actually some "dark matter" exerting a gravitational pull on all ranking schemes such that certain institutions or types of institutions (the Harvards, Oxfords, and Tsinghuas of the world) rise to the top regardless of the specific indicators and weightings used. A search for this "dark matter" certainly seems deserving of future research. One guess, however, is that "age of institution," "faculty size" and "per-student expenditure" are probably excellent candidates to be this "dark matter."

Rankings without League Tables: The CHE Approach

For most of this paper we have been describing league tables—that is, ranking systems that provide a single integrated score that allows an ordinal ranking of entire institutions. However, this is not the only possible approach to university rankings. There is, for instance, no intrinsic reason why indicators must focus solely on institutions; approaches that look at institutions at lower administrative levels (such as departments or faculties) are also possible. The *Guardian* and, as of 2006, *la Repubblica* both provide comprehensive departmental-level rankings *across entire universities* (that is to say, they provide separate rankings for each discipline), though they also synthesize the data upwards into institutional rankings, as the previous two sections have explored.

A different approach altogether is taken by the Centre for Higher Education Development (CHE) in Germany, which issues annual rankings jointly with a media partner (currently *Die Zeit,* formerly *Stern*). CHE conducts regular surveys of approximately 130,000 students and 16,000 faculty, covering nearly 250 higher education institutions. The student surveys are extensive and ask a number of questions about both student experiences and student satisfaction. The faculty survey is done in order to generate data for a special indicator known as the "insider's pick" (the survey asks professors to name the three institutions in their field of study that they would recommend to someone as the best places to study). The ranking also has a number of indicators that use independent sources of data. Roughly two-thirds of the indicators are survey-based (higher than any of the league tables listed in this study), and the remaining data points all come from third-party sources. The CHE rankings do not make use of university-sourced data.

The CHE ranking of German university departments differs from traditional league tables in two notable ways. First, as noted above, it does not weight or aggregate individual indicator scores. Each department's data on each indicator are allowed to stand independently, and no attempt is made to rank departments on an ordinal scale. CHE does this because it believes that it is at best meaningless (and at worst actively misleading) to combine widely disparate indicators into a single overall hierarchy.

There is a surprising level of agreement between ranking systems as to which universities in a given country are "the best."

This stance presents certain difficulties in presenting data in a printed format. Instead of a simple ordinal rank, all indicators must be shown for all institutions, which means that they are somewhat unwieldy and difficult to read. On the other hand, this stance has an enormous advantage when translated to the World Wide Web.[9]

Because CHE does not weight the ratings, it is possible for users themselves to in effect create their own weightings and rankings by selecting a restricted number of indicators and asking the Web site's database to provide comparative institutional information on that basis. In so doing, the CHE approach effectively cedes the power of defining "quality"—which, as we have seen, is one of the key roles arrogated by the authors of ranking schemes—to *consumers* of the ranking system (i.e., prospective university students and their parents or sponsors).

CHE's second unique point is that, even within each indicator, no attempt is made to assign ordinal ranks. Each institution's department in a given discipline is simply classified as being in the "top third," "middle third," and "bottom third" of all institutions with respect to that specific indicator. Schools within each of these three categories are considered qualitatively equal, on the grounds that for many indicators, ordinal rankings are spurious given the small difference in measurement.

Conclusions

Based on this survey of league tables, we can conclude a number of things. Perhaps most important, there are vast differences between university league tables in terms of what they measure, how they measure it, and how they implicitly define "quality."

Some of these differences appear to be geographic or cultural in nature. There is notable clustering of certain types of indicators and certain types of data sources. It is unclear whether this reflects genuine differences in opinion about the definition of what constitutes "quality" in universities or cross-national differences in the collection and availability of data. The lack of common indicators across countries explains why the large international league tables (*Shanghai Jiao Tong* and *THES*) are so reliant on measures of publication outputs and on reputational surveys (respectively), as they are the only indicators that do not rely on governments or institutions to first collect and process the data.

At the same time, despite major inconsistencies in the methodologies used to rank universities, there is a surprising level of agreement between ranking systems as to which universities in a given country are "the best." To the extent that different methodologies give differing opinions about the quality of an institution, the variance between observations grows as one moves down the ordinal rankings.

[9]Available at http://www.daad.de/deutschland/studium/hochschulranking/04690.en.html.

Finally, although the definition of "quality" is contested, league tables by definition impose a "one-size-fits-all" approach to the matter; this is precisely why they are so controversial. As the CHE approach shows, however, league tables are not the only way to approach rankings. Indeed, the spread of the World Wide Web provides collectors of institutional data with an opportunity to democratize rankings and put the power of ranking in the hands of the consumer by following an "any-size-fits-all" approach.

As Merisotis (2002) has noted, university rankings are here to stay. As imperfect as they are, they satisfy a public demand for transparency and information that institutions and governments have not been able to meet on their own. Moreover, as higher education becomes more costly for individuals and families, the demand for comparative information on universities will increase. As a means of delivering that information, however, league tables are only in their infancy, and all of them can clearly benefit from greater analysis of the assumptions implicit in their own schemes. This is particularly the case with respect to international league tables, which have a restricted range of possible indicators due to the lack of available cross-national comparative data. To the extent that international ranking schemes are taking on a quality assurance role in the growing international student market, this suggests that the global higher education community needs to begin to look at how best to collect and report data on institutions so as to permit thoughtful and responsible inter-institutional comparisons.

References

America's Best Colleges (Sklaroff, Sara, editor). 2005. Washington: U.S. News & World Report

Centrum Für Hochschulentwicklung (CHE, Germany: Berghoff, Sonia; Federkeil, Gero; Giebisch, Petra; Hachmeister, Cort-Denis; Siekermann, Meike; Müller-Böling, Detlef), 'Forschungsuniversitäten 2004,' No 62, 10 February 2005. Retrieved from http://www.daad.de/deutschland/hochschulen/hochschulranking/06543.en.html, 14 June 2005.

Daily Telegraph, 2003. League Tables: '*Who Really is Top of the League?*' Retrieved 04 March 2007 from http://www.telegraph.co.uk/education/main.jhtml?xml=%2Feducation%2F2003%2F06%2F27%2Ftefuni.xml. (Also accessed March-June 2005)

Dill, David D. & Soo, Maarja. 2004. Is There a Global Definition of Academic Quality? A Cross-National Analysis of University Ranking Systems. Chapel Hill, NC: Public Policy for Academic Quality background paper, University of North Carolina.

De Miguel, Jesús M.; Caïs, Jordi; and Vaquera, Elizabeth. 2001. *Excelencia – Calidad de las Universidades Españolas,* Madrid : Centro de Investigaciones Sociológicas, October 2001.

The *Financial Times. Financial Times Universities 2003*. 03 May 2003.

Finnie, Ross and Usher, Alex. 2005. *Measuring the Quality of Post-secondary Education: Concepts, Current Practices and a Strategic Plan*. Kingston, Ont.: Canadian Policy Research Networks.

Guangdong Institute of Management Science: Liu, Nian Cai & Liu, Li. 'University Rankings in China,' *Higher Education in Europe,* XXX (2, July 2005), 217-227.

The *Guardian* (Leach, Jimmy, editor). 2004. *The Guardian University Guide 2005: What to Study, Where to Go and How to Get There.* London: Atlantic Books.

The Hong Kong Institute of Vocational Education Network—Education18. 2006. Hong Kong's Best Universities, Eighth Edition. 'Ranking of Universities in Hong Kong in 2004'. Retrieved 26 August 2005 from http://www.education18.com/item04/0408/0408main_eng.htm.

Maclean's (Johnston, Ann Dowsett, editor). 2005. *Maclean's Guide to Canadian Universities '05.* Toronto: Rogers Publishing Limited.

Melbourne Institute Of Applied Economic And Social Research (Williams, Ross & Van Dyke, Nina). 2004. *The International Standing of Australian Universities.* Melbourne: University of Melbourne.

Merisotis, Jamie P. 2002. On the Ranking of Higher Education Institutions. *Higher Education in Europe,* XXVII (4), 361-363.

Netbig Chinese University Rankings 2003. Retrieved from http://rank2003.netbig.com/en/ 13 September 2005. Successive rankings also exist for 2004-2006.

Le Nouvel Observateur. 2003. 'Universités le Palmarès 2003,' *Le Nouvel Observateur,* 13 March 2003 (NB: Strictly speaking not a set of league tables, rather a descriptive/qualitative assessment of French universities.)

Perspektywy 2004 University Rankings. October 2004. Retrieved from www.perspektywy.pl/strona/images/zasady3.pdf on 10 July 2005.

La Repubblica, *La Grande Guida All'Università.* Retrieved from http://www.repubblica.it/speciale/2004/censis/index.html on 14 May 2005.

Shanghai Jiao Tong University—Academic Ranking of World Universities. Retrieved from http://ed.sjtu.edu.cn/ranking2005.htm on 21 April 2005.

The Times Good University Guide (O'Leary, John; Hindmarsh, Andrew; & Kingston, Bernard.) 2004. *The Times Good University Guide 2005.* Hammersmith, London: Times Newspapers Limited.

Times Higher Education Supplement—World University Rankings. Retrieved from http://www.thes.co.uk/worldrankings/ on 02 May 2005.

Usher, Alex, and Savino, Massimo. 2006 (January). A World of Difference: A Global Survey of University League Tables. Toronto, Ont.: Educational Policy Institute.

Van Dyke, Nina. Twenty Years of University Report Cards. *Higher Education in Europe,* XXX (2, July 2005), 103-125.

Wuhan University Centre for Science Evaluation. Liu, Nian Cai & Liu, Li. 'University Rankings in China,' *Higher Education in Europe,* XXX (2, July 2005), 217-227.

The Impact of Higher Education Rankings on Student Access, Choice, and Opportunity

By Marguerite Clarke[10]

Over the past two decades, rankings of higher education institutions and programs have emerged around the world. Produced by magazines, newspapers, research centers, and even governments, these rating systems play an increasingly important role as information tools for prospective students as well as marketing devices for institutions. The growing demand for rankings is fueled by several trends in higher education, including increasing participation rates, higher costs, and the view of students as consumers who expect value for money.

While there has been considerable research on the indicators and weights used to create these rankings (Clarke 2004; Dill & Soo 2005; Usher & Savino 2006; Van Dyke 2005), far less attention has been paid to their impact on students. This "consequential aspect" (Messick, 1994, p. 9) of the rankings phenomenon relates to both their intended and unintended effects on students, whether positive or negative.

As a step toward addressing this gap in the research literature, this chapter synthesizes some of the available evidence on rankings' impact in three important areas: student access, choice, and opportunity. While the focus is on findings for the United States, the chapter also makes comparisons to findings for other countries. It concludes with thoughts on the likely effects of the newest addition to the rankings scene—so-called "world" rankings that compare institutions or programs across national boundaries—on global outcomes in these areas.

Definitions and Brief Summary of Findings

The terms *access, choice*, and *opportunity* have multiple meanings, so it is work clarifying upfront how they will be used in this chapter. *Access* can be defined as "the process of enabling entry to higher education" (Harvey 2004). Given this definition, access can be examined in relation to the policies and procedures put in place by higher education institutions, or specific programs within those institutions, that directly impact entry for traditionally underrepresented student populations. This is particularly germane if one assumes that institutional behavior is affected by rankings, causing changes in institutional policies and procedures that impact educational opportunity.

The individual student-level process of choosing a college or graduate school has three main stages (Hossler & Gallagher 1987). In the first stage, the student becomes predisposed toward attending

[10]The author wishes to thank Kelvin Gregory and Jamil Salmi for their feedback on an earlier version of this chapter.

higher education as he or she develops educational and occupational goals. In the second stage, the student searches for information on schools and decides where to apply. In the third stage, the student decides to enroll in a particular school. This chapter focuses on the second and third stages in the *choice* process since this is where the impact of rankings is most likely to occur.

The long-term effects of access and choice after graduating—or, *opportunity*—can be seen in an individual's employment success and earnings. College graduates tend to have far higher average lifetime earnings than those with only a high school diploma (OECD 2006). Benefits are even greater for those with a professional or doctoral degree. It is important to recognize that there are other benefits from higher education. However, this chapter assesses opportunity in terms of employment and earnings outcomes for higher education graduates and examines how these outcomes relate to the characteristics (including the rank) of the institution or program attended.

The consequences for student access, choice, and opportunity vary, but tend to be particularly negative for low-income and minority students.

The findings presented here are based on a review of evidence from around the world on the impact of rankings on student access, choice, and opportunity.[11] The meaning (and meaningfulness) of the concepts of access, choice, and opportunity differ across countries. The extent to which rankings impact on each of these areas also differs depending upon the characteristics of a country's higher education system (e.g., structure, entry requirements, tuition policies) and the nature of the rankings themselves (e.g., source, purpose, indicators). Nonetheless, when data from various countries were reviewed, four themes emerged. First, while many countries have higher education rankings, data regarding their impact on students are available for only a few (in general, commercially-produced rankings in Western countries). Second, much of the available evidence on impact is anecdotal in nature; there are relatively few empirical studies. Third, when the available data are placed in the context of the broader literature on access, choice, and opportunity, it is evident that rankings are only one of many factors that have been found to impact student outcomes in these areas. Fourth, rankings—at least commercially-produced ones in Western countries—serve primarily to reinforce the effects of broader market-based and competitive forces in higher education. The consequences for student access, choice, and opportunity vary, but tend to be particularly negative for low-income and minority students. At least some of these adverse outcomes are related to the student selectivity indicators used in the rankings, and highlight the need for rankings that reward schools for how well they have educated students as opposed to how selective they have been in recruiting them.

Most of the evidence of rankings impact on student access, choice and opportunity comes from the United States. Because of this, much of this chapter focuses on U.S. findings. Where possible, however, it also draws comparisons to findings for other countries.

The U.S. Higher Education System

The American postsecondary education system comprises more than 5,000 colleges and universities, including public four-year colleges, private four-year nonprofit institutions, public two-year community colleges, and a variety of other types, such as private for-profit institutions

[11]Two main types of data were collected: published research and media reporting. The search was limited to English-language sources and to work published within the last twenty years.

that focus on vocational certificates. Over the years, various organizations have attempted to rank the institutions that make up this landscape (Salmi & Saroyan 2006), looking primarily at the four-year colleges that are selective in their admissions criteria. For example, the magazines *Money, Business Week,* and *U.S. News & World Report* have ranked subgroups of these schools according to their "value for money," student and recruiter satisfaction, and overall academic quality, respectively. The indicators used to create these rankings tend to be a mix of input and output measures that purport to reflect different facets of institutional or program quality. For example, *U.S. News* uses a weighted combination of seven broad categories of indicators—institutional reputation, student retention, faculty resources, student selectivity, financial resources, graduation-rate performance, and alumni-giving rate—to rank four-year colleges and universities according to their academic quality. Somewhat similar indicators are used for the magazine's rankings of graduate schools of business, education, engineering, law, and medicine. For instance, in the case of business schools, a weighted mix of three categories of indicators—program reputation, graduate placement success, and student selectivity—is used.

Most rankings of U.S. higher education institutions and programs are produced by magazines and newspapers. These organizations market the rankings as tools that will aid students in choosing among schools. The assumption seems to be that by providing prospective students with information on the relative quality of available options, they will choose a school in a way that optimizes their access and opportunity outcomes.

What is the Impact of Rankings on Student Access to this System?

The steady increase in U.S. higher education enrollment over the last five decades was facilitated by federal, state, and institutional policies and programs that aimed to remove economic, academic, and other barriers to access for particular student populations. Despite these efforts, the U.S. higher education system remains stratified by income and race/ethnicity, with low-income and minority (African-American and Hispanic) students concentrated in lower-price and less-selective institutions (Astin 2004). This stratification has been exacerbated in recent years by a relative decline in state expenditures for higher education, increased tuition costs for individual students, and the demise of race-based affirmative action in several states.

Rankings contribute to this increasing stratification of the U.S. higher education system by creating incentives for schools to recruit students who will be 'assets' in terms of maintaining or enhancing their position in the rankings. These incentives seem directly related to the student-selectivity indicators used in *U.S. News* and other rankings, including test scores for entering students, the proportion of entering students who graduated in the top ten percent of their high school class, and the percentage of applicants accepted.[12] In order to improve their performance on these indicators, schools engage in various strategic activities, including implementing early-decision programs, offering larger amounts of merit aid, and investing heavily in student consumption benefits.

Early-decision programs allow students to apply early to, and receive an early admissions decision from, their first-choice school. Machung (1998) notes that the *U.S. News* rankings have led to a growth in these programs because they allow schools to compete more effectively for academically

[12]These incentives are also related—albeit indirectly—to the student retention and graduation indicators used in the rankings, since low-income and minority students tend to 'perform' less well in these areas than other students.

high-achieving students. These programs have a negative impact on college access for low-income students, however, because if a student is admitted, they must withdraw their applications from all other schools. This means that they do not get a chance to compare the full range of financial aid offers they might have received had they opted to participate in the regular admissions process.[13] A related practice that schools use to improve their standing in the rankings is encouraging applications from students they don't plan to admit. This allows a school to reduce its acceptance rate, which is one of the indicators used in the *U.S. News* college (and graduate school) rankings (Dill & Soo 2005; Machung 1998).

Regarding merit aid, a recent report notes that U.S. colleges and universities are using enrollment management practices such as tuition discounting and financial aid leveraging to woo students with top test scores and grades (regardless of need) at the expense of needier students (Haycock, 2006). These practices, which include awarding scholarships and other non-need-based aid to reduce the cost of attendance, allow schools to 'purchase' talent that will make them look better in the *U.S. News* and other rankings. These practices effectively shut out low-income and minority students who are less likely to attend academically competitive high schools or earn top scores on standardized admissions tests. Vincent (2005) points out that while this trend began at private institutions, it is also evident at public universities; in fact, the percentage of financial aid that is merit based is actually higher at public universities (see also Baum & Lapovsky 2006). This is quite troubling considering that the traditional mission of public universities in the United States has been to expand access to higher education for underrepresented student groups.

In relation to student consumption benefits, a national study of U.S. colleges and universities (Brewer et al. 2002) found that many institutions were making extensive investments in dormitories, fiber optic computer networks, sporting and recreational facilities, and other infrastructure as a way to attract high-achieving students. The researchers suggest that this pursuit of prestige through image-enhancing facelifts (which do not necessarily improve the quality of the education that students receive) is reinforced by commercial rankings that use student inputs as a primary measure. One consequence of this investment is a reduction in the resources available for other activities, including those designed to recruit and retain students from traditionally underrepresented groups.

Similar strategies have been reported at the graduate level. For example, some business schools have been able to boost their position in the *U.S. News* and *Business Week* rankings by using financial aid to attract students with high test scores, and by creating two-tier MBA programs (students with low test scores are placed into part-time programs and those with high scores into the full-time program that 'counts' for the rankings) (Mufson 1999). Law schools also have used such tactics to boost their rankings (Sauder & Espeland 2006).

Given the detrimental effects of these activities on disadvantaged students, it seems clear that they contribute to the increasing stratification of the U.S. higher education system. Nonetheless, some observers argue that the *U.S. News* and other rankings have actually increased student access to an

[13]In September 2006, Harvard, Princeton, and the University of Virginia announced that they would end their early-admission programs (Farrell 2006). Harvard uses a version of early admission known as 'early action' (the offer of admission is non-binding on the student), while Princeton and the University of Virginia employ the early-decision model described above. Although early action is less restrictive than early decision, low-income students are still less likely to use it than wealthier students. Many doubt that these universities' actions will start a trend since they compete in a different league compared to most other schools and will likely attract the highest-achieving students no matter what their approach to admissions.

elite education. For example, Samuelson (2004) points out that because traditional top-tier schools lack space for all of the good students who apply, the institutions that enroll the overflow of high-quality applicants gain prestige. This raises the latter group of schools in the rankings. As Samuelson explains: "Schools that rise in the rankings acquire more prestige than falling schools lose. The result is more 'elite' schools—old elite plus new." What the author fails to consider, however, is that this dynamic does not necessarily result in greater access to elite schools for low-income and minority students.

Students who found rankings to be a very important factor in their choice of school were more likely to be high-achieving, from high-income families, and from families with college-educated parents.

Many countries with diversified higher education systems are similarly characterized by high enrollment rates combined with considerable stratification by income and minority/immigrant status. As in the United States, rankings that include student-selectivity indicators can contribute to this stratification by creating an additional incentive for schools to recruit the 'ablest' students. For example, *The Sunday Times* rankings of universities in the United Kingdom are compiled using a formula that rewards an institution's ability to attract high-achieving (in terms of their performance on national examinations) applicants above all else. Since university officials believe that applicants rely heavily upon these and other rankings to assess university quality, they focus on attracting more of these high-achieving students in order to enhance their position in future rankings (Rolfe 2003; Montgomery & Canaan 2004).

What is the Impact of Rankings on Student Choice in this System?

The institutions that comprise the U.S. higher education system together enroll approximately 18 million students. Research suggests that some mix of the following factors likely influenced these students' application and enrolment decisions: perceived academic quality and reputation of the institution in general and academic program in particular, entry requirements, location, tuition costs, financial aid availability, infrastructure, employment prospects for graduates, social life, advice of significant persons (e.g., family, friends, school personnel), and commercially-produced materials such as guidebooks and ranking publications (Hossler & Gallagher 1987; Kallio 1995; Lipman Hearne 2006; Perna 2006).

Research on the specific impact of rankings on these choice decisions has focused on two main areas: the types of students most likely to use rankings, and the effect that changes in a school's rank has on overall applications to, and enrollment in, that school. In relation to the first area, the evidence suggests that rankings do not play an important role in most students' application and enrollment decisions. For example, McDonough et al. (1998) report that only 11 percent of the 221,897 undergraduate students who responded to their survey saw commercial rankings as a very important factor in their choice of school; 60 percent found them not at all important (Lipman Hearne 2006). Students who found rankings to be a very important factor in their choice of school were more likely to be high-achieving, from high-income families, and from families with college-educated parents. They also were more likely to be Asian-American (or non-U.S. citizens), and to have intentions of getting doctoral, medical, or law degrees. Low-income and first-generation (i.e., children of parents with no higher education experience) college students were least likely to view the rankings as important. These findings are not surprising given that students of low socioeconomic status tend to enroll in community colleges and other non-selective institutions, which are generally not ranked in *U.S. News* or other systems.

In terms of the effect that changes in a school's rank has on overall application and enrolment decisions, Monks and Ehrenberg (1999) found that a less favorable *U.S. News* rank resulted in a declining applicant pool (at least for the selective, private, four-year institutions that they examined). In addition, a smaller percentage of admitted applicants matriculated, and the resulting entering class was of lower quality as measured by its average test scores (Meredith 2004). Similar findings have been reported for the *U.S. News* rankings of graduate schools of law (Sauder & Lancaster 2006; Wellen 2005) and business (Reese 1998). These outcomes are no doubt related to the type of students who find the rankings most useful—i.e., high-achieving students who wish to maximize the conversion capacity of their degree for further educational or occupational attainment (McDonough et al. 1996) and who use a school's rank as a guide to that capacity.

The research on student choice in other countries identifies a similar set of factors as influencing student choice decisions (Veloutsou et al. 2004). Research on the specific impact of rankings on these decisions is generally in line with the U.S. findings. For example, research from the United Kingdom (Carrico et al. 1997) and Germany (Federkeil 2002) confirms McDonough et al.'s (1998) finding that rankings are most often used by high-income and/or achievement-oriented students (Leach & Zepke 2006). Federkeil (2002) reports—similar to Monks and Ehrenberg (1999)—that a good result for a university in the Centre for Higher Education Development's rankings of German universities leads to an increase in applications to that university in the following year. However, Eccles (2002) concludes that there is no evidence to suggest that students react to changes in the position of an institution in the U.K. rankings.

What is the Impact of Rankings on Student Opportunity After Graduating from this System?

Every year, U.S. higher education institutions award more than a million bachelor's degrees, over half a million master's degrees, and a substantial number of doctorate and professional degrees. As mentioned earlier, bachelor's degree holders tend to have higher incomes and better jobs than those with only a high-school diploma. Master's, doctoral, and professional degree recipients do even better. What is of interest here is the fact that within each of these degree categories, employment and earnings outcomes vary considerably. Much of the variation has been traced to two factors: the demand for particular specializations, and the perceived status (also referred to as selectivity, reputation, or prestige) of the degree-granting institution (Thomas 2000). The strength of the relationship between each of these factors and employment and earnings outcomes is moderated by the state of the economy (Wu 2003).

Institutional status is of particular relevance to this discussion because it relates strongly to rankings. Ishida et al. (1997) found that the perceived status of the institution from which an individual received his or her undergraduate degree affected career advancement mainly at the beginning of that person's career; it did not seem to affect later career prospects. Among professional degrees, D'Aveni (1996) found that U.S. business schools were at a disadvantage in placing graduates if they lacked status in the national business community. At the doctoral level, studies have found a consistent positive correlation between the status of the departments where individuals received their Ph.D. degrees and the status of the departments where they obtained jobs, particularly their first jobs (Baldi 1995). This correlation holds for both the hard and social sciences and is independent of an individual's previous research performance.

What about the effects of status on earnings? In their review of the research, U.S. researchers Pascarella and Terenzini (2005) conclude that the impact of institutional selectivity on lifetime earnings is nonlinear. Only the most selective institutions may have an impact on earnings (Hoxby 1998). These authors also point out that the relationship depends upon a student's major field of study, which tends not to be controlled for in relevant studies. In addition, when studies control for the types of students who apply to more selective U.S. institutions, the earnings advantage of selective schools disappears. An important exception to the latter finding is children from low-income families, who were found to earn more if they attended selective colleges (Dale & Krueger 2002).

It is difficult to disentangle the specific impact of rankings on outcomes in these areas. One reason for this is that the terms status, rank, selectivity, reputation, and prestige are often used interchangeably in the research on this topic and authors do not always clarify whether they are using 'rank' to refer to general notions of status or to a school's position in a specific rankings publication.

The small amount of research that directly examines the effects of commercial rankings on employment and earnings outcomes suggests that these publications do have an impact, at least for business school graduates. For example, Reese (1998) notes that since business school graduates do not get a license, their school's position in the *U.S. News* and other rankings acts as a signal to employers; the higher the rank, the easier it is for a graduate of that school to gain access to certain companies, and to specific positions within those companies. In line with this theory, a recent study found that companies pay higher salaries to graduates of top-ranked American business schools even when they know that lower-ranked schools offer a better education (Rindova et al. 2005).

Students are aware that the rank of their school may affect their employment possibilities. Not surprisingly, students at less prestigious business schools have tried to increase the standing of their program in satisfaction-based rankings by sending back surprisingly upbeat surveys. There also are rumors of school officials coaching students on how to fill out these forms (Reese 1998).

Students are aware that the rank of their school may affect their employment possibilities.

Similar relationships between institutional status and employment/earnings outcomes have been found in other countries (Rolfe & Anderson 2003). The research is unclear as to the specific impact of rankings in these areas. In addition to the aforementioned confounding of terms, this lack of clarity is due to the fact that rankings tend to reflect traditional status hierarchies, which, in turn, are closely related to graduates' success on the job market. For example, in the U.K., *The Times Good University Guide* includes employment outcomes as one of its ranking indicators. These data show that graduates of the highest-ranked universities (Oxford, Cambridge, London School of Economics) have the best employment outcomes. However, since the highest ranked institutions also tend to be traditionally high-status schools, it is unclear which factor—if either—may be influencing employers.

In instances where rankings do not reflect traditional status hierarchies, it seems that traditional perceptions may hold sway. For example, in Nigeria, federal universities have long been viewed by employers as having higher standing than state universities. Recently, some Nigerian banks, ahead of their merger plans, began disengaging employees who were graduates of state universities (apparently as an image-enhancing exercise). This was done despite the fact that state universities

performed well against federal schools in a 2004 ranking of universities by the Nigerian Universities Commission (Olatunji 2005).

Institutional rankings also appear to have little or no sway on employers who hire graduates in very specialized and/or newly established program areas that are in high demand. For example, the aforementioned *Times Good University Guide* rankings reveal that while graduates of traditionally high-status institutions tend to do best overall in terms of employment prospects, graduates of some of the newer universities also perform well if they have specialized in an area highly regarded for particular professions.

Conclusion

Most critiques of rankings focus on the validity of the indicators and weights used (Clarke 2004). This chapter focused on a different aspect of the validity argument (Messick 1994)—rankings' consequences for students. The findings suggest that rankings—at least commercial ones in Western countries—impact student access, choice, and opportunity in ways that create unequal outcomes for different groups of students.[14] In particular, the rankings tend to most advantage high-income and high-achieving students and to most disadvantage minority students and those from low-income homes. At least some of these differential outcomes are related to the student selectivity indicators used in the *U.S. News* and other rankings, and highlight the need for rankings that reward schools for their relative success in educating students as opposed to their relative ability in recruiting already high-achieving ones.[15]

The rankings tend to most advantage high-income and high-achieving students and to most disadvantage minority students and those from low-income homes.

The last few years have seen a new addition to the rankings scene: so-called 'world' rankings that purport to be lists of the top universities or programs in the world. These rankings reflect the fact that education increasingly operates in a global environment, and also provide this global market with its "performance measure" (Maslen 2005). The best-known examples are the *Times Higher Education Supplement* (THES) World University Rankings, and the Academic Ranking of World Universities produced by the Institute of Higher Education at Shanghai Jiao Tong University (SJT). There also are rankings of the top programs in a specific area, of which the *Financial Times* ranking of the top MBA programs in the world is probably the best known. The THES rankings are primarily based on peer opinion/reputation, while the SJT rankings have a strong emphasis on research performance, and the *Financial Times* rankings rely heavily on indicators of postgraduate career success (e.g., employment, earnings, and promotions).

Evidence is just starting to emerge as to the impact of these rankings on student access, choice, and opportunity. Taken together, the data suggest that the outcomes may be somewhat similar to those seen for commercial rankings in the U.S and other Western countries.

Regarding access, the potential impacts appear mixed. For example, these global rankings seem to be opening up new possibilities for students, alerting them to countries where they can obtain

[14]Since most of the available data focus on commercially produced rankings in Western countries, it would be useful to collect data on the impact of other types of rankings—e.g., those used by governments as quality assurance or funding mechanisms (Bunnag 2006; Salmi & Saroyan 2006).
[15]See the Berlin Principles on Ranking of Higher Education Institutions in the appendix to this monograph.

a high-quality education for a reasonable price (Davie 2006). The lists also are being used by some governments as a 'value for money' indicator. For example, the Mongolian government is considering making available funding for study abroad only to those students who are admitted to a university that appears in one of the world rankings (Mongolian Ministry of Education, Culture, and Science official, personal communication, December 10, 2005). At the same time, universities aspiring to higher positions in these rankings are starting to compete more aggressively in the international marketplace for the highest-performing students at the expense of domestic undergraduates (Braude & Sharma 2005). Institutions already doing well in the rankings have, on occasion, used their position to justify charging high tuition (Merola 2006), while those doing poorly (e.g., Irish universities) have used their lackluster performance to call for the introduction of student fees (McConnell 2005). These rankings also appear to be putting pressure on some national systems to free at least one university to pursue a meritocratic approach to education (without regard to quotas or affirmative action) so that the country will have representation among the world's leading universities ("Comparisons Matter" 2005). These types of institutions may focus on research and graduate student education, admit students only based on merit, and use their rank to justify the increasing costs of the education they deliver.

Rankings have been linked to changes in national and cross-national application patterns as achievement-oriented students seek the globally top-ranked program in their area.

Regarding choice, the rankings have been linked to changes in national and cross-national application patterns as achievement-oriented students seek the globally top-ranked program in their area (Davie 2006). Such applicant shifts—which are similar to those seen in response to the *U.S. News* rankings—have the potential to affect a country's share of the international market for foreign students (Maslen 2005).

In terms of opportunities after graduation, some observers note that in a global economy that draws on an increasingly international labor market, employability will come to depend more on the global status or rank of the university conferring the degree.[16] One example of this is a British Treasury decision to waive regular visa requirements for graduates of the top 50 MBA programs in the world (based primarily on rankings by the Financial Times), so that they can more easily work in Britain (Cohen 2005). It is worth noting that this example pertains to business school graduates, the same segment of the U.S. higher education population whose employment prospects seemed most affected by the standing of their school in the U.S. News and other national rankings.

Messick (1989) notes that "To appraise how well [a measuring instrument] does its job, we must inquire whether the potential and actual social consequences of…interpretation and use are not only supportive of the intended…purposes, but at the same time are *consistent with other social values*" (p. 8, emphasis added). The above findings—both for national and world rankings—suggest that the impact of rankings on students is more consistent with some social values (e.g., meritocracy, competition) than with others (e.g., access, equity). Whether these outcomes are deemed appropriate depends, in large part, on the value system that comes to characterize the global higher education environment.

[16]Montgomery & Canaan (2004) also point out that a growing linkage seems to be developing between each institution's location within this hierarchy of institutions and the kinds of jobs for which they are preparing students.

References

Astin, A. W. (2004). The Declining "Equity" of American Higher Education. *The Review of Higher Education, 27*(3), 321-341.

Baldi, S. (1995). Prestige Determinants of First Academic Job for New Sociology Ph.D.s 1985-1992. *The Sociological Quarterly, 36*(4), 777-789.

Baum, S., & Lapovsky, L. (2006). *Tuition Discounting: Not Just a Private College Practice.* New York: College Board. Retrieved September 13, 2006, from http://www.collegeboard.com/prod_downloads/press/tuition-discounting.pdf

Braude, J., & Sharma, Y. (2005, November 5). Patten Backs More China Students for Oxford University. Fee-Paying Students from India and U.S. Urged Along with Cut in Local Enrolment. *South China Morning Post,* p. 2.

Brewer, D., Gates, S. M., & Goldman, C. A. (2002). *In Pursuit of Prestige: Strategy and Competition in U.S. Higher Education.* New Brunswick, NJ: Transaction Press.

Bunnag, S. (2006, January 15). Universities to be Rated on Academic Standards. *Bangkok Post.*

Carrico, C. S., Hogan, S. M., Dyson, R. G., & Athanassopoulos, A. D. (1997). Data Envelope Analysis and University Selection. *The Journal of the Operational Research Society, 48*(12), 1163-1177.

Clarke, M. (2004). Weighing Things Up: A Closer Look at *U.S. News & World Report's* Ranking Formulas. *College and University, 79*(3), 3-9.

Cohen, D. (2005, February 16). Salaries Put Paid to Rankings. *The Australian,* p. 44.

Comparisons Matter. (2005, November 11). *New Straits Times.*

Dale, S., & Krueger, A. (2002). Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables. *The Quarterly Journal of Economics, 117*(4), 1491-1527.

D'Aveni, R. A. (1996). A Multiple-Constituency, Status-Based Approach to Interorganizational Mobility of Faculty and Input-Output Competition Among Top Business Schools. *Organization Science, 7*(2), 166-189.

Davie, S. (2006, April 14). Varsity Applications at Record High—Surge Due to Larger Cohort, More Flexible Admission Policy. *The Straits Times.*

Dill, D. D., & Soo, M. (2005). Academic Quality, League Tables, and Public Policy: A Cross-National Analysis of University Ranking Systems. *Higher Education, 49*(4), 495-533.

Eccles, C. (2002). The Use of University Rankings in the United Kingdom. *Higher Education in Europe, 27*(4), 423-432.

Farrell, E. (2006, September 26). U. of Virginia Announces Plans to End Early-Decision Admissions. *The Chronicle of Higher Education.* Retrieved September 26, 2006, from http://chronicle.com/daily/2006/09/2006092604n.htm

Federkeil, G. (2002). Some Aspects of Ranking Methodology—The CHE Ranking of German Universities. *Higher Education in Europe, 27*(4), 389-397.

Harvey, L. (2004). *Analytic Quality Glossary.* Retrieved September 10, 2006, from the Quality Research International Web site: http://www.qualityresearchinternational.com/glossary/access.htm

Haycock, K. (2006). *Promise Abandoned: How Policy Choices and Institutional Practices Restrict College Opportunities.* Washington, D.C.: The Education Trust.

Hossler, D., & Gallagher, K. S. (1987). Studying College Choice: A Three-Phase Model and the Implications for Policy Makers. *College and University, 2,* 207-221.

Hoxby, C. (1998). *The Return to Attending a More Selective College: 1960 to the Present.* Retrieved September 6, 2006, from Harvard University Department of Economics Faculty Web site: http://www.economics.harvard.edu/faculty/hoxby/papers/whole.pdf

Ishida, H., Spilerman, S., & Su, K. (1997). Educational Credentials and Promotion Chances in Japanese and American Organizations. *American Sociological Review, 62*(6), 866-882.

Kallio, R. E. (1995). Factors Influencing the College Choice Decisions of Graduate Students. *Research in Higher Education, 36*(1), 109-124.

Leach, L., & Zepke, N. (2006). *Student Decision-Making by Prospective Tertiary Students. A Review of Existing New Zealand and Overseas Literature.* Wellington, New Zealand: Ministry of Education.

Lipman Hearne. (2006). *High-Achieving Seniors and the College Decision.* Retrieved September 8, 2006, from http://www.lipmanhearne.com/teenstudy/LHI-teenstudy.pdf

Machung, A. (1998, July/August). What's at Stake: College Rankings and the News Media. *Change,* 12-16.

Maslen, G. (2005, October 22). Shanghai University Sets Global Standard on Research. *South China Morning Post,* p. 2.

McConnell, D. (2005, August 21). No Show in Top 200 is 'Disaster,' Say Colleges. *The Sunday Independent.*

McDonough, P. M., Antonio, A. L., & Horvat, E. M. (1996, August). *College Choice as Capital Conversion and Investment: A New Model.* Paper presented at the meeting of the American Sociological Association, New York.

McDonough, P., Antonio, A. L., Walpole, M., & Perez, L. X. (1998). College Rankings: Democratized College Knowledge for Whom? *Research in Higher Education, 39*(5), 513-537.

Meredith, M. (2004). Why Do Universities Compete in the Ratings Game? An Empirical Analysis of the Effects of the *U.S. News & World Report* College Rankings. *Research in Higher Education, 45*(5), 443-461.

Merola, A. (2006, April 23). Uni with Class. Expensive? It's a Matter of Degrees. *Sunday Mail,* p. 020.

Messick, S. (1989). Meaning and Values in Test Validation: The Science and Ethics of Assessment. *Educational Researcher, 18*(2), 5-11.

Messick, S. (1994, June). *Validity of Psychological Assessment: Validation of Inferences from Persons' Responses and Performances as Scientific Inquiry into Score Meaning.* Keynote address presented at the Conference on Contemporary Psychological Assessment, Stockholm, Sweden.

Monks, J., & Ehrenberg, R. (1999). *The Impact of U.S. News & World Report College Rankings on Admissions Outcomes and Pricing Policies at Selective Private Institutions* (NBER Working Paper 7227). Cambridge, MA: National Bureau of Economic Research.

Montgomery, L. & Canaan, J. (2004). Conceptualizing Higher Education Students as Social Actors in a Globalizing World: A Special Issue. *International Journal of Qualitative Studies in Education, 17*(6), 739-748.

Mufson, S. (1999, March 14). Rankings All-Important to GWU. Business School's Drive Controversial. *The Washington Post,* p. A01.

Olatunji, B. (2005, November 8). State Varsity Graduates Now Untouchables? *This Day.*

Organisation for Economic Co-operation and Development (OECD). (2006, September). *Education at a Glance 2006: OED Indications.* Paris: Organization for Economic Cooperation and Development.

Pascarella, E. T., & Terenzini, P. T. (2005). *How College Affects Students: A Third Generation of Research* (2nd ed.). San Francisco: Jossey-Bass.

Perna, L. (2006). Studying College Access and Choice: A Proposed Conceptual Model. In J. C. Smart (Ed.), *Higher Education: Handbook of Theory and Research: Vol. XXI* (pp. 99-157). Netherlands: Springer.

Reese, J. (1998, September). The Brouhaha Over Rankings. *Stanford Business.* Retrieved September 8, 2006, from http://www.gsb.stanford.edu/community/bmag/sbsm9809/feature_brouhaha.html

Rindova, V., Williamson, I., & Petkova, A. (2005). Being Good or Being Known: An Empirical Examination of the Dimensions, Antecedents, and Consequences of Organizational Reputation. *The Academy of Management Journal, 48*(6), 1033-1049.

Rolfe, H. (2003). University Strategy in An Age of Uncertainty: The Effect of Higher Education Funding on Old and New Universities. *Higher Education Quarterly, 57*(1), 24-47.

Rolfe, H., & Anderson, T. (2003). A Firm Choice: Law Firm's Preferences in the Recruitment of Trainee Solicitors. *International Journal of the Legal Profession, 10*(3), 315-334.

Salmi, J., & Saroyan, A. (2006, May). *League Tables as Policy Instruments: The Political Economy of Accountability in Tertiary Education.* Paper presented at the Second Meeting of the International Rankings Expert Group, Berlin, Germany.

Samuelson, R. J. (2004). In Praise of Rankings. *Newsweek.* Retrieved September 8, 2006, from http://msnbc.msn.com/id/5626554/site/newsweek

Sauder, M., & Espeland, W. N. (2006). Strength in Numbers? The Advantages of Multiple Rankings. *Indiana Law Journal,* 81(205), 205-227.

Sauder, M., & Lancaster, R. (2006). Do Rankings Matter? The Effects of *U.S. News & World Report* Rankings on the Admissions Process of Law Schools. *Law and Society Review, 40*(1), 105-134.

Thomas, S. (2000). Deferred Costs and Economic Returns to College Quality, Major, and Academic Performance: An Analysis of Recent Graduates in Baccalaureate & Beyond. *Research in Higher Education, 41,* 281-313.

Usher, A., & Savino, M. (2006). *A World of Difference: A Global Survey of University League Tables.* Toronto, ON: Educational Policy Institute.

Van Dyke, N. (2005). Twenty Years of University Report Cards. *Higher Education in Europe, 30*(2), 103-125.

Veloutsou, C., Lewis, J. W., & Paton, R. A. (2004). University Selection: Information Requirements and Importance. *International Journal of Educational Management, 18*(3), 160-171.

Vincent, F. (2005, May 29). No Merit in these Scholarships. *The Washington Post,* p. B1.

Wellen, A. (2005, July 31). The $8.78 Million Maneuver. *The New York Times,* p.18.

Wu, A. (2003, October 26). Personal Business. A Tight Job Market Dampens Ivy League Hopes. *The New York Times,* p. 8.

Next Steps

The three papers included in this monograph reinforce the notion that rankings are coming to be recognized globally as a major force in higher education. Among the most important impact of ranking systems is how they affect what can be characterized as a three-dimensional accountability marketplace for higher education. The first part is the system of peer-based *institutional accreditation,* which is now grappling with many issues, including the global phenomenon of education and ensuring quality across borders. The second is the strategy of *governmental oversight* of higher education institutions, either through direct accreditation or through regulatory control of institutional processes and transactions. *Rankings* by magazines, newspapers, and other organizations—while often not happily welcomed by colleges and universities—have clearly emerged as the final major dimension of this accountability marketplace. As each of the authors has made clear, whether or not colleges and universities agree with the various ranking systems is largely irrelevant. Ranking systems are popular and clearly here to stay.

A critical question that has emerged from this work and from the Berlin Principles discussion is how dialogue, research, and dissemination about ranking systems will be maintained at the global level, and how different ranking systems will be assessed in terms of their coherence with the Berlin Principles. With that in mind, it will be important to consider ways in which this global work on ranking can continue and expand. This enhanced agenda should include:

- Conducting, commissioning, and encouraging research that assesses ranking systems and contributes to the development of new knowledge about how rankings impact quality improvement in higher education;
- Convening ranking organizations and analysts to review the development of new ranking systems, and consider modification or amendment with regard to the scope and methodologies of ranking that is consistent with the Berlin Principles;
- Facilitating and mediating dialogue among higher education institutions, ranking organizations, and analysts; and
- Assessing the coherence of various rankings with the standards of good practice outlined in the Berlin Principles.

This represents an enormous set of challenges that, nevertheless, must be addressed if we are to achieve the goals of a system of continuous improvement and refinement of these increasingly essential market-based accountability mechanisms in higher education.

Moving the international rankings research, convenings, and standards agendas forward has direct implications for rankings and accountability discussions in the United States. Within American higher education, the three dimensions of accountability discussed above are often

convergent concerns for policymakers, institutional administrators and practitioners, researchers, and consumers of higher education. Having better information about how other countries and international organizations are addressing rankings and their intersections with institutional accreditation and governmental oversight can provide important new avenues for addressing the growing national concern about how we define student success, how we make student learning outcomes transparent, and how institutions are held accountable for what they do.

In addition, constantly evolving American rankings will continue to provide important information for the international dialogue. As aptly highlighted here by Sanoff, U.S. ranking systems are regularly influenced by broader discussions of accountability, student success measurements, and data transparency and availability. National developments such as proposed alternative models for institutional rankings and comparisons, the emergence of new data sources like the National Survey of Student Engagement and the call for more robust data collection by federal agencies, and potential changes to national accountability and accreditation requirements may inform practice in other countries grappling with similar concerns.

Therefore, U.S. engagement in the international work must be twofold. First, American researchers, rankings providers, and higher education practitioners and policymakers must model the continuous improvement of rankings systems called for at the international level. To accomplish that goal, practitioners can continue to generate research and dialogue about U.S. rankings systems in an effort to refine current rankings, debate alternatives, and discuss new ways of defining and documenting student and institutional success. Second, we must also continue and broaden our participation in international rankings dialogues, sharing lessons from our national work with international counterparts and leveraging the rankings experience of other countries to inform rankings in our own country.

The proposed international agenda laid out in this report would provide new opportunities for individual countries like the United States to contribute to and learn from rankings systems worldwide. While important for countries with well-established rankings systems, such access would particularly benefit countries just beginning rankings work and those with fewer resources to engage in this work on their own. The agenda can also situate and extend important ongoing dialogues about issues such as access and success of less-advantaged students, the relationship between the world-class university movement and rankings, and the role developed nations play in supporting rankings development in less developed nations.

International rankings work provides exciting opportunities for cross-national collaboration, shared learning, innovative approaches to rankings, and common assessments of current and emerging ranking systems. This proposed international agenda provides the means for making this work more robust, accessible, and transparent and for enhancing the multinational collaborations and dialogues begun in recent years.

Appendix

The Berlin Principles on Ranking of Higher Education Institutions

Rankings and league tables of higher education institutions and programs are a global phenomenon. They serve many purposes: they respond to demands from consumers for easily interpretable information on the standing of higher education institutions; they stimulate competition among them; they provide some of the rationale for allocation of funds; and they help differentiate among different types of institutions and different programs and disciplines. In addition, when correctly understood and interpreted, they contribute to the definition of "quality" of higher education institutions within a particular country, complementing the rigorous work conducted in the context of quality assessment and review performed by public and independent accrediting agencies. This is why rankings of higher education institutions have become part of the framework of national accountability and quality assurance processes, and why more nations are likely to see the development of rankings in the future. Given this trend, it is important that those producing rankings and league tables hold themselves accountable for quality in their own data collection, methodology, and dissemination.

In view of the above, IREG was founded in 2004 by the UNESCO European Centre for Higher Education (UNESCO-CEPES) in Bucharest and IHEP. It is upon this initiative that IREG's second meeting (Berlin, 18 to 20 May, 2006) has been convened to consider a set of principles of quality and good practice in higher education rankings—**the Berlin Principles on Ranking of Higher Education Institutions.**

It is expected that this initiative has set a framework for the elaboration and dissemination of rankings—whether they are national, regional, or global in scope—that ultimately will lead to a system of continuous improvement and refinement of the methodologies used to conduct these rankings. Given the heterogeneity of methodologies of rankings, these principles for good ranking practice will be useful for the improvement and evaluation of rankings.

Rankings and league tables should:

A) Purposes and Goals of Rankings

1. *Be one of a number of diverse approaches to the assessment of higher education inputs, processes, and outputs.* Rankings can provide comparative information and improved understanding of higher education, but should not be the main method for assessing

what higher education is and does. Rankings provide a market-based perspective that can complement the work of government, accrediting authorities, and independent review agencies.

2. *Be clear about their purpose and their target groups.* Rankings have to be designed with due regard to their purpose. Indicators designed to meet a particular objective or to inform one target group may not be adequate for different purposes or target groups.

3. *Recognize the diversity of institutions and take the different missions and goals of institutions into account.* Quality measures for research-oriented institutions, for example, are quite different from those that are appropriate for institutions that provide broad access to underserved communities. Institutions that are being ranked and the experts that inform the ranking process should be consulted often.

4. *Provide clarity about the range of information sources for rankings and the messages each source generates.* The relevance of ranking results depends on the audiences receiving the information and the sources of that information (such as databases, students, professors, employers). Good practice would be to combine the different perspectives provided by those sources in order to get a more complete view of each higher education institution included in the ranking.

5. *Specify the linguistic, cultural, economic, and historical contexts of the educational systems being ranked.* International rankings in particular should be aware of possible biases and be precise about their objective. Not all nations or systems share the same values and beliefs about what constitutes "quality" in tertiary institutions, and ranking systems should not be devised to force such comparisons.

B) Design and Weighting of Indicators

6. *Be transparent regarding the methodology used for creating the rankings.* The choice of methods used to prepare rankings should be clear and unambiguous. This transparency should include the calculation of indicators as well as the origin of data.

7. *Choose indicators according to their relevance and validity.* The choice of data should be grounded in recognition of the ability of each measure to represent quality and academic and institutional strengths, and not availability of data. Be clear about why measures were included and what they are meant to represent.

8. *Measure outcomes in preference to inputs whenever possible.* Data on inputs are relevant as they reflect the general condition of a given establishment and are more frequently available. Measures of outcomes provide a more accurate assessment of the standing and/or quality of a given institution or program, and compilers of rankings should ensure that an appropriate balance is achieved.

9. *Make the weights assigned to different indicators (if used) prominent and limit changes to them.* Changes in weights make it difficult for consumers to discern whether an institution's or program's status changed in the rankings due to an inherent difference or due to a methodological change.

C) Collection and Processing of Data

10. *Pay due attention to ethical standards and the good practice recommendations articulated in these Principles.* In order to assure the credibility of each ranking, those responsible for collecting and using data and undertaking onsite visits should be as objective and impartial as possible.

11. *Use audited and verifiable data whenever possible.* Such data have several advantages, including the fact that they have been accepted by institutions and that they are comparable and compatible across institutions.

12. *Include data that are collected with proper procedures for scientific data collection.* Data collected from an unrepresentative or skewed subset of students, faculty, or other parties may not accurately represent an institution or program and should be excluded.

13. *Apply measures of quality assurance to ranking processes themselves.* These processes should take note of the expertise that is being applied to evaluate institutions and use this knowledge to evaluate the ranking itself. Rankings should be learning systems continuously utilizing this expertise to develop methodology.

14. *Apply organizational measures that enhance the credibility of rankings.* These measures could include advisory or even supervisory bodies, preferably with some international participation.

D) Presentation of Ranking Results

15. *Provide consumers with a clear understanding of all of the factors used to develop a ranking, and offer them a choice in how rankings are displayed.* This way, the users of rankings would have a better understanding of the indicators that are used to rank institutions or programs. In addition, they should have some opportunity to make their own decisions about how these indicators should be weighted.

16. *Be compiled in a way that eliminates or reduces errors in original data, and be organized and published in a way that errors and faults can be corrected.* Institutions and the public should be informed about errors that have occurred.

1320 19th Street, NW, Suite 400
Washington, DC 20036
www.ihep.org