



# STEM and the History of the University Ranking Movement: Contextualizing Trends in Methodologies and Criteria

*William R. Stevenson III*

Wherever universities exist, and for as long as they have existed, there have been debates over which schools are the most prestigious or which can boast the highest quality of learning. Nevertheless, it has only been over the past one hundred years that such conjecturing has given way to data-driven rankings. In the beginning, rankings included only a limited number of American universities and served primarily as a source of reference for a small group of scholars. More recently, enabled by technological advances, rankings have incorporated bigger data and used increasingly complex equations to rank institutions from around the world. The results have changed the culture of higher learning. Today, rankings not only affect prospective students, but they also impact university agenda and governmental policy. In particular, they have led to increased emphasis on research-intensive STEM fields, often at the

---

W. R. Stevenson III (✉)

Department of Education and Culture, Doshisha University, Kyoto, Japan  
e-mail: wstevens@mail.doshisha.ac.jp

© The Author(s) 2018

J. N. Hawkins et al. (eds.), *New Directions of STEM Research and Learning in the World Ranking Movement*, International and Development Education,  
[https://doi.org/10.1007/978-3-319-98666-1\\_2](https://doi.org/10.1007/978-3-319-98666-1_2)

expense of the social sciences and humanities. Yet, while recent rankings use more data and wield greater influence across the globe, they have changed little in terms of methodology, remaining captive to the same criteria that characterized such lists from the start. The following provides an overview of past rankings, focusing on the role of STEM fields in particular, with the goal of establishing a deeper and more contextualized understanding of the ranking movement and its current impact on higher education.

### MEN OF SCIENCE AND THEIR UNIVERSITIES

The modern ranking movement began at the turn of the twentieth century, appearing in connection with the publication of various articles and books that focused on the backgrounds of prominent individuals. To use the title of Alick Maclean's brief study, the common goal was to understand *Where We Get Our Best Men* (1900) and, in some cases, women. While authors such as Havelock Ellis in *A Study of British Genius* (1904) were hesitant to link education to success, concluding that great individuals "owe a remarkably small proportion of their learning to the established machinery of instruction" (p. 148), others were eager to connect intelligence or social prominence to learning. Of particular importance to the birth of university rankings was John Leonard's *Who's Who in America*. The original 1899 edition includes 8602 names of notable living Americans and opens with an "educational statistics" section in which Leonard argues that education is among the "especially prominent" characteristics shared by the successful men and women referenced in his study. Leonard, however, did not attempt to measure the connection between education and success. And, despite the urging of an unnamed "scientific man," for reasons of "time and space" he declined to include lists that would show which institutions had produced "the most eminent men," suggesting instead that the readers do their own calculations (p. xii).

Within a few years, psychologist and science advocate James McKeen Cattell (1860–1944) took up the challenge. He began by compiling a reference work entitled *American Men of Science*, which grew out of an earlier list created for the Carnegie Institution of Washington. The volume itself does not rank schools, but Cattell (1906, p. v) claimed that it was the first work to provide a "fairly complete survey of the scientific activity of a country at a given period," which could be "even more

useful in academic circles than ... *Who's Who in America*.” The first edition of the work includes biographical sketches of over four thousand scientists from twelve designated fields. Cattell and his assistants selected the four thousand from roughly ten thousand questionnaires sent to persons believed to have “contributed to the advancement of pure science” based on their belonging to scientific societies, their contribution to scientific research and writing, or their inclusion in other lists such as *Who's Who in America* (pp. v–vi). Cattell then added a star to one thousand of the entries—a quarter of the listed scientists—whose work was thought to be “the most important” (p. vii). He selected these individuals by having ten leading scientists from each of the twelve fields arrange the names of persons within their field “in order of merit” (p. vii).

While Cattell’s work is significant for its novel approach to determining the status of individual scientists, what makes *American Men of Science* of particular interest to the current rankings movement are two papers that he wrote in the process of creating the larger work. First, in a 1903 article for the *American Journal of Psychology*, Cattell took a select group of two hundred American psychologists and did a publication count to compare their influence with that of Europeans. In Cattell’s words, “to compare our productivity with that of other nations, I have counted up the first thousand references in the index of the twenty-five volumes of the *Zeitschrift fur Psychologie*” (p. 327) and concluded that “it appears that each of our psychologists has on the average made a contribution of some importance only once in two or three years” (p. 328). Apart from highlighting an Atlantic divide that no longer exists, the study is notable for being one of the earliest uses of bibliometrics to establish academic hierarchy.

The second study appeared in *Science*, the official journal of the American Association for the Advancement of Science and a publication that Cattell personally owned.<sup>1</sup> In brief, the study took the one thousand distinguished scientists from the 1906 volume and calculated their number *and* list placement to compile a ranking of institutions based on “scientific strength.” The top five results were Harvard, followed by Chicago, Columbia, John Hopkins, and Yale. Cattell made no pretense of having considered anything more than the production of scientific knowledge. He wrote that while “a university may conceivably have a

<sup>1</sup>Cattell also included the results of the study in the second edition (1910) of *American Men of Science*.

department consisting of men of moderate scientific standing, but of personal distinction and superior teaching ability ... such men belong to the past rather than to the present generation.” After all, though admittedly conjecture, Cattell argued that “scientific men of ability and character will be investigators, and there is a high correlation between these traits and teaching skill” (Cattell 1910, pp. 684–685).

Current research shows this conclusion to be wrong. In their oft-cited study, Marsh and Hattie (2002) indicate that “teaching effectiveness and research productivity are nearly uncorrelated,” and that “research performance does not provide a surrogate measure of teaching effectiveness, nor do measures of teaching effectiveness provide an indication of research productivity” (p. 635). Nevertheless, Cattell believed that scientific research should not only be the primary purpose of the university, but that it was the foundation for all of industrial civilization. “Science and its applications,” he wrote, “should be the chief concern of a democratic nation that would preserve its democracy and advance the freedom and the welfare of its people” (Cattell 1922, p. 278). While few today would argue that research production should be the sole measure of a university, Cattell’s studies pioneered several aspects of university assessment that continue to reverberate: his work was the first large study based on informed opinion gathered through a questionnaire, it was the first to focus on STEM field research, it was one of the first to use some form of bibliometrics and, finally, it was the first to be readily accessible. Its value, in Cattell’s (1910, p. 688) own words, was to “show the advantage of statistics over general impressions.”

### GRADUATE SCHOOL RANKINGS

Despite the originality of Cattell’s study, he never updated his rankings. It would be another scientist, albeit one who had long abandoned his work as a chemist for administrative duties, to take the next step in university assessment. Raymond M. Hughes (1873–1958), president of Miami University, set out to evaluate and rank graduate programs from across the United States in the early 1920s (Hughes 1925). His goal was to produce a reference guide for Miami students looking to attend graduate school. Turning first to his university’s faculty, he asked them to create a list of “distinguished national scholars” in 20 designated fields. He then sent each listed scholar a questionnaire, which became the basis of his assessment. He initially presented his study in a speech before the

members of the Association of American Colleges and, according to Cartter (1966), “stirred up considerable interest, and no little criticism,” but “presumably had an impact on that student generation” (p. 5).

Perhaps Hughes’ harshest critic was future university ranker Hayward Keniston (1883–1970). Writing a quarter century later, Keniston (1959) described Hughes’ effort as being dependent on “highly subjective impressions” that were subject to the “halo of past prestige.” The end result, according to Keniston, was a ranking of no “real validity” apart from providing a “fairly close approximation to what informed people think about the standing of the departments in each of the fields” (p. 117). Keniston, a well-known scholar of Romance languages, came out of retirement in the late 1950s to work as a consultant to the University of Pennsylvania, heading their effort to update Hughes’ study. The goal was to determine the position of Pennsylvania’s graduate programs relative to those at other leading schools.

Despite his criticisms of Hughes, Keniston followed a similar approach. The only significant difference is that he limited his survey to department chairs who, in his opinion, “by virtue of their office ... know what is going on at other institutions” (p. 117). He asked them to rank graduate programs based on a combination of faculty reputation and perceptions of program quality. The results were then compiled to provide twenty-four departmental rankings. He then merged the lists to rank graduate programs in four general areas (biological sciences, humanities, physical sciences, and social sciences), and finally combined the data to produce an institution-wide ranking. Even while critical of Hughes’ earlier study, Keniston chose to list his findings alongside the 1925 rankings to assess quality gains and losses over the previous quarter century. He concluded that several universities, primarily state schools, had noticeably improved while others, such as Chicago, had lost much of their status.

Less than a decade later, economist and vice-president of the American Council of Education, Allan M. Cartter (1922–1976), led a new study that set out to update earlier rankings and assess on a far broader scale the graduate programs of all “major universities” in the United States. His primary criticisms of earlier rankings, and of Keniston in particular, was that they relied too heavily on department chairs: a demographic that in Cartter’s view tended to be older, more conservative, outdated in perception, and not necessarily the most informed or distinguished scholars in their field. He also argued that both Hughes

and Keniston had failed “to separate measures of faculty quality from measures of educational quality” (Cartter 1966, p. 6). A valid survey, he thought, needed to make a clear distinction between the “scholarly reputation” of faculty and the value of a program in terms of the students’ “educational experience” (p. 9).

Cartter’s approach was to survey 4008 junior scholars, senior scholars, and department chairs representing 29 fields of study from 106 institutions, leading to an assessment of 1663 doctoral programs. His questionnaires distinguished between “quality of faculty” and “effectiveness of graduate program,” resulting in a more nuanced ranking that, in his own opinion, was “as reliable a guide as one can devise in attempting to measure the elusive attribute of quality” (p. 9). Cartter was keenly aware of the subjective nature of university rankings. In fact, he opens his study by explaining that “in the final analysis the national reputation of a department or an institution is nothing more than an aggregation of individual opinions” (p. viii). As such, Cartter chose to limit his assessment to programs and, almost in passing, to five “general areas of study.” He refused to combine scores to create a university-wide ranking, writing that such an effort would be arbitrary as it would involve “some judgement about how the various fields of study should be weighted” (p. 106).

Cartter’s ranking was similar to those of Hughes and Keniston in its dependence on informed opinion. Where it differed, besides scale and its use of a more nuanced questionnaire, is that Cartter chose to go deeper in his analysis, choosing four of the twenty-nine fields (economics, English, political science, and physics) for a more detailed study that included bibliometrics. The method echoed the earlier efforts of Cattell, but rather than count the number of references made to scholarly works, Cartter initiated a method that is still used today. He selected major journals from each of the four fields (the number of journals varying with each field) and, looking at a four-year period, counted the number of articles, shorter communications, and book reviews published by the faculty of each institution. In addition, he tallied the number of books, textbooks, and edited volumes reviewed in the same journals and assigned each type of publication a designated weight. Taking into consideration the unique character of each discipline, Cartter and his team selected different weight ratios for each of the four fields. Unlike many ranking systems today that use bibliometrics as a core or sole indicator, assessing universities based on rates of quality publication, Cartter only used bibliometrics to examine the correlation between the results

of his ranking and the production of scholarship (pp. 78–105). In the field of economics, for example, there was a clear correlation between the strengths of a program and publication rates. In the field of English, however, Cartter found the correlation to be far less pronounced, with the faculty of weaker programs often producing work at a rate that would put them on par with scholars from higher ranked schools (pp. 80, 88).

According to Gourman (1977), “the academic community’s response to the Cartter report was overwhelming,” inspiring “widespread comment and critique” (p. 7). Within five years, 26,000 copies were distributed, which was followed by a second study headed by the American Council on Education in 1970 that used the same basic methods (Roose and Andersen 1970). In addition, even though Cartter refused to turn his assessment of graduate programs into an institution-wide ranking, others quickly did so using the results of his study. Horace Magoun (1907–1991), for example, published an institution-wide ranking within the same year, justifying his article on the grounds that “such syntheses are of value today because of the extent to which activities related to graduate education have come to determine the intellectual and economic well-being of the communities and regions in which graduate schools are situated.” Continuing, he writes, “In our contemporary society, many extra-mural groups and agencies are interested in the over-all standings of universities and their divisions” (Magoun 1966, p. 483). Magoun does not specify any “groups” or “agencies,” but the implication is clear: long before the current proliferation of university rankings, long before schools began to aggressively look for ways to improve their international standing, the link between “economic well-being” and the “overall standing of universities” was being established.

## THE RANKING EXPLOSION

While Cartter’s study reached a far broader audience than that of Cattell, Hughes, or Keniston, all four rankings were produced by academics for academics. It was only the involvement of media corporations and major publishing houses that eventually resulted in university assessment going mainstream. The *Chicago Tribune* was possibly the first newspaper to publish a university ranking, listing the best ten undergraduate programs in a widely discussed piece by Chesly Manly. Although based on a survey of prominent educators, the criteria for “best” was largely

left to respondents (Stuit 1960, p. 375). Meanwhile, the postwar surge in college enrollment rates created a market for college guidebooks. Among the earliest was the College Entrance Examination Board's *Annual Handbook*, but it was Barron's *Profiles of American Colleges* (1964 to present) that began to rank the universities according to categories of "most competitive" to "noncompetitive." James Cass and Max Birnbaum's *Comparative Guide to American Colleges* (Harper and Row, 1964–1991) and *Peterson's Annual Guide to Undergraduate Study* (first published in 1970, and currently titled *Peterson's Four-Year Colleges*) followed a similar pattern. While the publications surely encouraged prospective students to think of universities in hierarchal terms, their assessments were admittedly subjective. Cass and Birnbaum, for example, wrote that their categories were "not a measure of the overall quality of colleges, which are far too complex to be ranked by simple statistical data" (1989, 14 ed., p. x).

In 1983, *U.S. News & World Report* began publishing a biennial review of schools that contained both guidebook elements and a straightforward institutional ranking. Although the methods were initially dubious, by the end of the decade the company was producing an annual standalone issue, "America's Best Colleges," that used various combinations of survey data along with previously unreleased information provided by the institutions. The magazine was not the first to combine criteria, but it pioneered the use of "inside" data that gave its results an aura of authority. According to Usher (2017), universities "could still criticize the use of survey data in the rankings, ... [but] the debate was no longer really about whether multi-indicator rankings were measuring quality or not; the debate accepted that assumption, and moved on to the question of whether the methodology was correct."

Using their marketing know-how, *U.S. News & World Report* turned the study of university assessment into a lucrative business. With everyday Americans hungry to learn which universities topped each year's lists, traditional powerhouses, such as Stanford, Cornell, and Yale began to "play the rankings game" by looking for ways to improve their status (Machung 1998). Rankings were no longer a mere measurement of university quality, they were now shaping the direction of higher education policy.

The "America's Best Colleges" approach quickly became the standard, and a model for companies and organizations across Europe, Asia, and



the Americas. Over the ensuing years, *Maclean's* in Canada, *The Times* and *The Guardian* in England, *Asahi* and *Yomiuri* in Japan, *Der Spiegel* in Germany, and others created their own national rankings, experimenting with different combinations of indexes. And, with top universities becoming increasingly global in scope, it was only a matter of time before the rankings became global in scale. By the early twenty-first century, international student recruitment had become widespread, which according to Harvey (2008), served as the prime incentive for creating international university rankings. Shanghai's Jiao Tong University led the way in 2003, resulting in the current Academic Ranking of World Universities (ARWU; est. 2009). QS World University Rankings (est. 2004) came next, followed by the Dutch CWTS Leiden Ranking (est. 2007), the Thomas Reuters' *Times Higher Education (THE)* rankings (est. 2010), the Saudi Arabian Center for World University Rankings (CWUR) rankings, and, among others, *U.S. News & World Report* Best Global Universities rankings (est. 2014).

The global rankings market is now flooded with competitors, and their growing influence has led to a sense of unease among many in academia. Much of the worry is based on the tendency of rankings to assess what some believe to be a limited picture of higher education. In particular, critics are quick to point at assessments based on citation indexing services, namely Elsevier's Scopus and Thomson Reuter's Web of Science, that favor STEM fields with their high rates of publication. Those in the humanities in particular, "see this phenomenon as a colonization of their domain through a system that has mainly been applied (and probably can only be applied) in the positive sciences" (Loobuyck 2009, p. 209).

While online indexing services are a recent development, university rankings have never gone unchallenged. In 1910, during the same year that Cattell published the first analytical ranking of colleges, the American Association of Universities asked historian-turned-administrator Kendrick Babcock to assess the quality of higher education. His rankings, leaked to the press, caused such an uproar (particularly among those affiliated with schools that failed to make an impression) that President William Taft—and later Woodrow Wilson—banned their publication (Webster 1986). Similarly, in 1957, when the *Chicago Tribune* became the first newspaper to produce a list of best undergraduate programs, Northwestern University's school paper complained that the listing had "done a lot of harm" and may have "damage[d] them materially" (*The Stanford Daily*, 1957). Finally, in a critique of his own

pioneering work, Allan Cartter wrote that, “No single index ... nor any combination of measures is sufficient to estimate adequately the true worth of an educational institution” (Cartter 1966, p. 4). While debates continue as to which measures are the most reliable or relevant, Cartter’s observations remain as pertinent today as they were 50 years ago: “In an operational sense,” he wrote, “quality *is* someone’s subjective assessment, for there is no way to objectively measuring what is in essence an attribute of value” (Cartter 1966, p. 4).

## CONCLUSION

For over a century universities have been assessed and ranked according to outcome-oriented methodologies, including the use of bibliometrics and reputation surveys. STEM fields have been core to these indexes from the start. In fact, the first impression of a hundred-year survey of university rankings is that remarkably little has changed. Of the limited number of new approaches, the most notable was likely OECD’s Assessment of Higher Education Learning Outcomes (AHELO). Conducted in 2012 as a “feasibility study,” it looked to scientifically assess “what students in higher education know and can do upon graduation” (<http://www.oecd.org>). Seen as a threat to more traditional forms of assessment, the OECD announced in 2015 that the project would be placed on hold (Morgan 2015). In short, for generations, there has been surprising consistency in the general methodologies used to evaluate schools. It is only in the details that differences between past and current systems begin to emerge, and it is only in distinct emphases that competing systems of ranking produce their various results.

What has changed, and what makes the current world rankings movement cause for concern is its seemingly uninhibited growth in scale and influence. World rankings today receive nearly universal coverage, eliciting responses from every corner of the globe. The University of Nairobi’s “development studies” program appears on a QS ranking and the school celebrates for being included in the “top 100 universities across the globe” (QS Rankings 2016); Shanghai loads the top of their rankings with American universities and contributes to a surge of international students that brings more than thirty-five billion dollars annually into the US economy; London’s *THE* ranking demotes the University of Malaya and results in politicians calling for the resignation

of the Minister of Education (*The Malaysian Times* 2014); meanwhile, the Russian government, like many others, recently established a massive grant aimed at getting five national universities ranked among the top 100 in the world. Today, the implications of university rankings go far beyond past systems of assessment. No longer just a reference for prospective students or administrators, rankings are now center to a high-stakes competition, an “academic arms race” (Rhoads et al. 2014) that is fought on a global scale.

Despite the widespread belief that world rankings have too much influence, or that existing rankings compromise the educational role of universities by placing too much emphasis on STEM field research, there is no immediate alternative to the current system. Nevertheless, in agreement with Altbach, as “the inevitable logic of globalization make them a permanent part of the 21st-century ... [t]he challenge is to understand their nuances, problems, uses—and misuses” (2012, p. 31). The tendency to lump all schools together, irrespective of contexts, emphases, or stated purposes, and the impulse to place more value on statistically friendly research production than on the more opaque measurements of quality instruction are particularly worrisome. All of these trends—compounded by the economic implications of a knowledge economy—have led to an increased emphasis on STEM fields. Nevertheless, as shown, the first significant ranking of universities was, in fact, a STEM field ranking. For better or worse, the developed world has been moving toward a STEM-oriented future for generations and, as is becoming more evident, the rate of acceleration has outpaced the ability of non-STEM fields to adapt. This was the dilemma that John Plumb described in his 1964 *The Crisis in the Humanities*, but which a half-century of developments have failed to remedy. Academic fields that are underrepresented in university rankings and, by implication, have less to offer in a knowledge economy, are today struggling to maintain their place in higher learning; in some cases, they are frantically looking for students to justify their continued existence. Cartter began his 1966 study by ranking programs within the humanities, leaving STEM fields for the end. Since his writing, enrollment rates for the humanities has dropped by half (*Harvard Magazine* 2013). Meanwhile, the most recent Nobel prize for literature—the only Noble marked for a non-STEM field—went to a guitar-wielding folksinger who aptly prophesied that “the times, they are a-changing.”

## REFERENCES

- “Addressing a Decline in Humanities Enrollment.” 2013. *Harvard Magazine*, June 6. <http://harvardmagazine.com>.
- Altbach, Philip G. 2012. “The Globalization of College and University Rankings.” *Change: The Magazine of Higher Learning* 44 (1): 26–31.
- Carter, Allan M. 1966. *An Assessment of Quality in Graduate Education*. Washington, DC: American Council on Education.
- Cass, James, and Max Birhbaum. 1989. *Peterson’s Guide to Four-Year Colleges*, 14th ed. Princeton, NJ: Peterson’s Guides.
- Cattell, J. Mckeen. 1903. “Statistics of American Psychologists.” *American Journal of Psychology* 14: 310–328.
- , ed. 1906. *American Men of Science: A Biographical Directory*. New York: The Science Press.
- . 1910. “A Further Statistical Study of American Men of Science.” *Science* 32 (828): 672–688.
- . 1922. “The Organization of Scientific Men.” *The Scientific Monthly* 14 (6): 568–578.
- Gourman, Jack. 1977. *The Gourman Report: A Rating of Undergraduate Programs in American and International Universities*. Los Angeles: National Education Standards.
- Harvey, Lee. 2008. “Rankings of Higher Education Institutions: A Critical Review.” *Quality in Higher Education* 14 (3): 187–207.
- Hughes, Raymond M. 1925. *A Study of the Graduate Schools of America*. Oxford, OH: Miami University.
- Keniston, Hayward. 1959. *Graduate Study and Research in the Arts and Sciences at the University of Pennsylvania*. Philadelphia: University of Pennsylvania Press.
- Leonard, John W., ed. 1899. *Who’s Who in America: A Biographical Dictionary of Living Men and Women of the United States, 1899–1900*. Chicago: A. N. Marquis & Company.
- Loobuyck, Patrick. 2009. “What Kind of University Rankings Do We Want?” *Ethical Perspectives* 16 (2): 207–224.
- Machung, Anne. 1998. “Playing the Rankings Game.” *Change* 30 (4): 12–16.
- Magoun, Herbert W. 1966. “The Carter Report on Quality in Graduate Education: Institutional and Divisional Standings Compiled from the Report.” *Journal of Higher Education* 387 (9): 481–492.
- Marsh, Herbert W., and John Hattie. 2002. “The Relation Between Research Productivity and Teaching Effectiveness.” *Journal of Higher Education* 73 (5): 603–641.
- Morgan, John. 2015. “OECD’s Ahelo Will Fail to Launch.” September 21. <https://www.timeshighereducation.com>.

- “‘Muhyiddin Is the One Should Resign’, Slams Kit Siang.” 2014. *The Malaysian Times*, June 30. <http://www.themalaysiantimes.com>.
- “QS Rankings Place University of Nairobi Among Top 100 Universities Across the Globe.” March 23, 2016. <http://mombasa.uonbi.ac.ke>.
- Rhoads, Robert A., Shuai Li, and Lauren Ilano. 2014. “The Global Quest to Build World-Class Universities: Toward a Social Justice Agenda.” *New Directions for Higher Education*. 168 (Winter): 27–39.
- Roose, Kenneth D., and Charles J. Andersen. 1970. *A Rating of Graduate Programs*. Washington, DC: American Council on Education.
- Stuit, Dewey B. 1960. “Evaluations of Institutions and Programs.” *Review of Educational Research* 30 (4): 371–384.
- “Tribune’s College Ratings Cause Controversy.” 1957. *The Stanford Daily*, 131 (52) (May 7).
- Usher, Alex. 2017. “A Short Global History of Rankings.” In *Global Rankings and the Geopolitics of Higher Education*, edited by Ellen Hazelkorn (pp. 23–53). London and New York: Routledge.
- Webster, David S. 1986. *Academic Quality Rankings of American Colleges and Universities*. Springfield, IL: C. C. Thomas.