

The Berlin Principles on Ranking Higher Education Institutions: Limitations, legitimacy, and value conflict

Abstract:

University rankings have been widely criticized and examined in terms of the environment they create for universities. In this paper I reverse the question by examining how ranking organizations have responded to criticisms. I contrast ranking values and evaluation with those practiced by academic communities. I argue that the business of ranking higher education institutions is not one that lends itself to isomorphism with scholarly values and evaluation and that this dissonance creates reputational risk for ranking organizations. I argue that such risk caused global ranking organizations to create the Berlin Principles on Ranking Higher Education Institutions, which I also demonstrate are decoupled from actual ranking practices. I argue that the Berlin Principles can be best regarded as a legitimizing practice to institutionalize rankings and symbolically align them with academic values and systems of evaluation in the face of criticism. Finally, I argue that despite dissonance between ranking and academic evaluation there is still enough similarity that choosing to adopt rankings as a strategy to distinguish one's institution can be regarded as a legitimate option for universities.

Keywords:

University rankings, performance measurement, ranking ethics, values, evaluation, university governance, cultural economy.

Research Ethics

The research upon which this paper is based was approved by a University Research Ethics Board. All participants were provided with informed consent.

Author:

Gary RS Barron
Department of Sociology, Faculty of Arts
15-7 HM Tory Building, Edmonton, AB, Canada T6G 2H4
email: grbarron@ualberta.ca
phone: (780) 492-5234
fax: (780) 492-7196

To cite:

Barron, Gary RS. 2016. "The Berlin Principles on Ranking Higher Education Institutions: Limitations, legitimacy, and value conflict". *Higher Education*, Online First, pp.1-17.
DOI: 10.1007/s10734-016-0022-z

Introduction

Much of university life involves valuing and evaluating people, their work, and relative performance. Perhaps the most maligned of these practices is university rankings. Studies of university rankings have primarily argued that rankings produce effects on higher education through creating risk to institutional reputations and disciplinary surveillance (Power, Scheytt, Soin, and Sahlin, 2009; Sauder and Espeland, 2009). In this paper I invert the question by examining how ranking organizations have responded to critics. Using The Berlin Principles on Ranking Higher Education Institutions (see: <http://ireg-observatory.org/en/index.php/berlin-principles>) as its focus, this paper discusses how the principles work as standards in the production of legitimacy for ranking organizations. I demonstrate how the Berlin Principles are decoupled (Orton and Weick, 1990) from actual ranking practices—that is—the practices are very different from what the principles propose. I argue the Berlin Principles can be best regarded as a legitimizing practice to institutionalize rankings and symbolically align them with academic values and systems of evaluation. I also argue that the ranking business does not lend itself to alignment with scholarly values and evaluation, that this dissonance creates reputational risk for ranking organizations requiring work at legitimization.

University systems of evaluation are based on a logic of heterarchy that is dissonant with the hierarchies promoted by rankings. Heterarchy involves a combination of lateral accountability, organizational heterogeneity, and diversity (Stark, 2009). Academic culture does promote stratification according to relative status, but these heterarchies allow such statuses to exist while promoting individuals based on local and discipline specific values. Scholars also hold scientific values of transparency, methodological and explanatory rigor, and peer review. Peer review is also a means for professors to move up through the academic ladder. For example, promotion from assistant professor to full professor involves review of one's work in individually contextualized ways. Increasing reputation and prestige is necessary for success in academia, but how such review is done is much different from

how rankings are produced. Rankings—in their composition and visualizations—are characterized by competition in creating a zero sum game where progress up the hierarchy is only possible by the decline of others, distinction (where there is much similarity), simplification (in their approach to complex objects and use for decision making), exclusion (of alternative values, cultures, and models of organizing), and obfuscation (of business practices, and research methods). As a result, while ranking organizations produce reputational risk for higher education institutions, ranker's create reputational risk for themselves by objectifying a community with values and evaluation practices that are dissonant from their own. For universities and ranking organizations the risks are in regard to their claims to be legitimate institutions. For universities, to not be ranked or to fall significantly in the rank may harm reputation based on observers questioning the quality of their education programs or research opportunities. For ranking organizations the risk is in regard to their claim to be a legitimate form of assessment.

I demonstrate that decoupling is due to the fact that ranking work is distributed across diverse organizations (there are many different types of universities with many types of organizational missions and which exist in many types of political-economic-societal contexts which collect and report data in unstandardized ways) which are incentivized by rankings to engage in unprincipled data practices (see Sauder and Espeland, 2009), but also because ranking technology does not lend itself to tight coupling with the Berlin Principles. Finally, I argue that despite dissonance, the similarity in some ranker and university values (such as their interest in distinction), and the symbolic capital offered by rankings makes them a viable strategic option for university leaders. That is, rankings seem to align with academic values of peer review and prestige so when university's engage in strategic planning they seem to speak to their values. However, the processes of how those values are realized, the means of valuing, are far from aligned with how academics assess one another. The dissonance between rankings and academia is in the differences between their means of valuing.

This topic is not only a matter of importance for higher education institutions, but also for studies of organizations and society more broadly. In particular, Lamont (2012) has called for better understanding of conditions that sustain heterarchies under present conditions characterized by, “a context in which definitions of worth that are not based on market performance tend to lose their relevance and in which market fundamentalism is exercising strong homogenizing pressures on collective identities and shared definitions of what defines worthy life” (p.210). Higher education is a unique field for such studies because universities are defined—at least in part—by their rare ability to incorporate diversity (Smelser, 2013). University rankings commensurate the values and categories used in universities to think about and engage in academic work and transform them into something else. How we value is an important social and cultural form and it is inherently political (Espeland, 1998). Transforming values and categories of thought has consequences for what can or cannot be debated, what sort of work can or cannot be valued in higher education, and can fundamentally change what an organization is able to do and how it can do it.

My discussion proceeds with a brief overview of university rankings. I then elucidate the 16 Berlin Principles and describe their complexities. As my discussion unfolds, I provide insights into rankings, the information infrastructure (Bowker and Star, 1999) upon which rankings rely, and how rankings and related practices have implications for higher education systems.

Method

This paper is the product of research on rankings, performance metrics, and governance in universities that used insights from institutional ethnography (Smith, 2006) and science and technology studies (Law and Hassard, 1999) to trace how local values and practices of evaluation within universities are connected with extra-local activities to produce university rankings. This broader project involved interviews with 61 professors, deans, support staff, and people working at ranking organizations;

observations at three rankings-related conferences; involvement in rating a university on its sustainability performance; involvement on a number of university governing committees; attendance at several workshops on performance metrics and related practices in universities; and reading hundreds of news media articles and other documents related to rankings. I have illustrated categories of people I interviewed in Table 1.

Sampling was undertaken largely through convenience, but based on institutional ethnography's (Smith, 2006) interpretation of how organizations and institutions operate. Institutional ethnography posits that institutions have formed because of text mediated action. An actor will orient toward policies, procedures and job descriptions to go about their work day. During that work day they will often depend on others to get their job done. By asking actors about who they work with and how they go about doing their work a researcher can effectively trace out relationships between organizational roles and processes, thereby identifying new people to interview and contribute additional information. I began by acquiring publicly available contact information for university administrators and employees of ranking organizations. Once I made contact with an individual I would find further people to interview through our conversation about organizational work, this approach is common to institutional ethnographers (Smith, 2006). By asking people who work in organizations about their jobs how they get information and who they engage with to get their jobs done I would learn about other people internal or external to their organizations. For example, I would ask a dean where they get information for their annual reports, they would name someone in an institutional analysis unit, then I would contact that person and ask who they submit data to at ranking organizations, follow up on that contact and so on. In this way I was not only able to learn about connections between individuals within universities that are necessary to each completing their work, I was also able to do so with regard to individuals beyond specific organizational boundaries. An additional benefit to this approach is that I was able to verify information that any research participant

provided by asking others across the organization about it. For example, if a dean in business told me they use journal rankings to determine their professor's merit, tenure, or promotion, I could ask another dean or department chair if that was the case because these organizational actors all take part in interdisciplinary review of each other's work. If I was curious about a process one organizational actor had explained to me, I would often ask other participants for their view of the process to see if the two accounts matched. I further triangulated what I learned from interviews with official organizational documentation such as policies, procedures, collective bargaining agreements, public reports, promotional material and so on. By checking interviews for consistency with other interviews and other documentation I was able to further validate the information that I received.

Table 1: Research participant's by category in order of organizational hierarchy with frequency

Participant	Frequency
Board Chair	1
President	1
Provost	1
Vice-President	6
Dean	14
Department Chair	9
Professors	2
University Institutional Analysis	7
Recruiter (national/international)	6
Marketing Communications	2
Other University Staff	4
Ranking Editor	3
Ranking Sales/other	5
Total	61

For this specific discussion, I use my accumulated understanding to critique the Berlin Principles. Following Lynch (1993), I use examples where I find them, rather than deploy any systematic method of choosing. Accordingly, I frequently use examples from the rankings with which I am most familiar. To be sure, there are many other ranking systems than the ones discussed here, but rankings are a specific technology, they rely on quantification and reduction to order things into a hierarchy. As such, all rankings share similar limitations. For example, a car is a specific technology, it is good for some things and not good for others. A car can get you from point A to point B relatively quickly, but if point B is across an ocean the car is not the technology that you want. Regardless of who you are or where you are in the world the car's limitations will affect you in the same ways. This is not to say that cars do not have local contingencies, just that despite those contingencies all people will face the same limitations presented to them by a car unless they make significant modifications to it, effectively making it a new technology. Is a car with wings and which flies still a car? This example is merely illustrative, I'm not suggesting that rankings can necessarily be modified to be made into another technology altogether and it is not the purpose of the discussion herein to consider such work.

Further, standardization promoted through rules like the Berlin Principles aim to reduce local contingencies. For example, if every government across the globe implemented the same traffic laws the ways in which people use cars around the world become more predictable. Again, local contingencies would exist, but if one were to observe traffic in Canada and compare it to the UK one would make much more similar observations than one would in the absence of such standards. Rankings are a specific technology and the Berlin Principles seek to standardize practices around them. As such, my discussion will be applicable to most, if not all university rankings the reader may encounter, but this does not mean that people think about or enact rankings exactly the same across locations. My observations were largely made in Canada, and so the discussion primarily considers academic values and assessment within a North American context. Yet again, academics across the

globe share values and practices in common and expect particular standards. This is why if one were to attend an academic conference anywhere in the world one can predict reasonably well how it will proceed. With these limitations in mind, I articulate how rankings operate as a common technology then move into my critique of the Berlin Principles.

Rankings

Rankings are a technology of visibility and simplification. In general, the purpose is to order objects to allow the rapid identification and location of each item within the list. In university rankings institutions are ordered from most excellent to least excellent. This is because before rankings are made exclusionary criteria are applied so that only the “best”—for example, up to 500 of the 15,000-20,000 universities that exist in the world—are selected to be ranked. Then, universities are ordered on a continuum based on particular criteria that are believed to identify excellence or world-class status.

Ranking processes, such as creating numerical representations and placing them on an ordinal scale, is the fabric of many university operations. In particular, academe is the source of a culture of assessment with interest in ensuring that what “is and ought will be brought into proper alignment and the best candidate will be declared Number One” (Strathern, 1997:307). Rankings made by non-academic organizations have adopted academic assessment practices and made them their own. Their ostensible purpose is to aid decision making. Yet, rankings create substantial profits for their producers and may increase costs for universities, since participation often requires universities to hire staff to build databases, gather and standardize data to grow repositories, and produce reports.

The three largest and most commonly followed rankings are the The Times Higher Education (THE) World University Rankings, Academic Ranking of World Universities (ARWU), and Quacuarelli Symonds (QS) World University Rankings. THE is a journalistic magazine based in London, UK and its rankings measure university performance across all of the core characteristics of

“world class” universities such as teaching, research, knowledge transfer, and international outlook (THE, 2014). The ARWU—also known as the Shanghai Rankings—began as an academic exercise at the Centre for World Class Universities of Shanghai Jiao Tong University in China, and aimed to benchmark Chinese universities against the rest of the world. ARWU's measures center on markers of research excellence, primarily the number of Nobel and other international prizes a university has won (Liu and Cheng, 2005). Both THE and QS partner with Elsevier in order to use citation data from its Scopus database as measures of research excellence. THE also contracts with Elsevier to conduct and analyze the reputation surveys that account for up to 34.5% of the weight in their ranking. The QS ranking makes use of survey data for 60% of its ranking weight: 20% is based on employer's view of recruits from universities, and 40% is based on surveys of professors and university administrators. Each ranking organization also asks participating universities to submit data through an online form.

Rankings are controversial and are criticized for using poor research methods, homogenizing higher education by promoting one university archetype, emphasizing science and technology over arts and humanities, being biased toward English speaking universities, and engendering inequality and exclusion (Amsler and Bolsmann, 2012; Hazelkorn, 2011; West, 2009).

In Berlin, during spring of 2006, a group of people prominently involved in producing university rankings met to develop a set of principles for their work. This group was named the International Rankings Expert Group (IREG). IREG created 16 standards for ranking, the “Berlin Principles on Ranking Higher Education Institutions,” and these are the foundation for IREG's rankings audit. These principles have become a touchstone for rankings related practices and guide how many well-known organizations produce their rankings (Marope, Wells, and Hazelkorn, 2013). As of March 2015, Perspektywy University Rankings in Poland, Center for Higher Education Ranking in Germany, and the QS World University Rankings have all been audited by IREG and can now put the “IREG Approved” logo on their promotional materials (IREG, 2015). As we shall see, there are inherent

complications in putting the principles into practice.

Academic evaluation

Academic values and standards vary by discipline. In business, peer-reviewed journal articles published in highly ranked journals are particularly desirable. In economics, oncology, sociology, extension, and the arts, plenary addresses at major conferences, practice guidelines, books, community engagement, and dramatic productions are regarded as important outputs. In my interviews with deans, department chairs, faculty and staff, I learned that in evaluative practices around hiring, tenure and promotion, academics try to evaluate one another not by imposing their individual values and interests on their colleagues, but by assessing their colleagues' work according to the values and practices appropriate to their own disciplinary and sub-disciplinary criteria. Academic evaluation is heavily contextualized and performance is made transparent through written summaries provided by those being evaluated, letters of reference from esteemed colleagues, spoken advocacy by department chairs and deans. This process is heteronomous in that it allows diverse values and outputs to prosper. These findings are similar to those articulated by Lamont (2009) in her study of peer review to determine access to prestigious academic fellowships, and described by Sayer (2016) in his critique of the REF in the UK. Lamont (2009) also demonstrates that academics value competition as a means of distinguishing the most outstanding colleagues by bestowing symbolic awards upon them. University rankings appear to fit well with the values of peer review, distinction, and competition, but we shall see that their processes are not well aligned with transparency, heteronomy, lateral accountability, or the rigor that many academics strive for in their daily work. In what follows, I examine the 16 Berlin Principles and demonstrate how they are decoupled from actual ranking practices.

Berlin Principles: The dream and the reality

Purposes and goals of rankings

1. Be one of a number of diverse approaches to the assessment of higher education inputs, processes, and outputs.

The first principle responds to concern that people become overly focused on rank and ignore other methods of evaluation. At one meeting I attended to inform a university's faculty about a sustainability ratings initiative on campus the presenters noted more than once that “it's not supposed to be a ranking system, but we make it into one”, and the creators of ranking systems have also acknowledged that even with additional information “most people will only look at the table” (Walker, 2004). When a range of organizations are assigned a grade in relation to their performance people identify the highest score, the second highest score, the third, and so on, until an unofficial cognitive ranking is established in their minds. Where assessment with any degree of consequence exists, people will rank. Where rankings are a part of broader assessment, people will focus on rank because it is quick, easy to interpret, and illustrates a relative position that holds hope for security or supports rhetorical assertions of one's excellence. Ranking is a means of constructing a relative identity and creating distinctions where none exist, are relatively inconsequential, or are unclear. That is not to say that other information will not come into account, university administrators I spoke with often used rankings with other metrics in order to to make comparisons and inform decisions, but they did not typically probe weightings, or sub-measures that compose the rankings.

2. Be clear about their purpose and their target groups.

This principle derives from ranker comments that people misuse rankings, that they must “give information that is more relevant to the needs of users such as universities, students, and policymakers...” (Nigavekar, 2012). Clarity about target groups and the purpose of a ranking might

make sense for smaller scale, local, rankings, but defies the logic of comprehensive global or world class university rankings. Such rankings have a broad audience and what the audience believes or does with the ranking will not be determined by merely stating the proper purpose and audience. Moreover, uses of rankings and their data change over time. At one conference I attended, a renowned rankings organization representative proclaimed that rankings have become a geopolitical tool. Some governments have also started to develop immigration policies based on university rankings (Hazelkorn, 2011). I learned from university international recruitment staff that rankings are used for determining what universities students will be allowed to attend under national study-abroad scholarship programs, and that special administrative units within universities are being funded to attract and support fully sponsored international students (non-sponsored students are left to fend for themselves). Global rankings quickly become accepted as valid criteria to be used among audiences for many purposes, including national identity formation and sorting students.

3. Recognize the diversity of institutions and take the different missions and goals of institutions into account.

The need to recognize diversity in mission and goal is a rebuttal to the critique that rankings have a homogenizing effect by promoting a single model of what a university should be (Marginson and Wende, 2007). Following this principle results in more types of rankings, or the adaptation of rankings to present them in multiple categories. For instance, Global rankings offer a break down by university discipline, rankings of universities under 50 years old, or University by region (Baty, 2014; Times Higher Education, 2014; Quacerelli Symonds, 2013). Such divisions slice up the higher education sector and produce new sorts of hierarchies that did not previously exist while maintaining existing hierarchies. Rather than having one hierarchy where all universities could be recognized, there are now many hierarchies that are hierarchically organized.

Recognizing diversity in higher education institutions serves the purposes of rankers. It fits with their marketing logic and business. In practice such recognition is a form of niche standardization, where individuality and uniqueness is not recognized so much as new groups are made for purposes of further scrutiny and marketing (Timmermans and Epstien, 2010). Such recognition leads to more ranking-related products to sell and stratifies the post-secondary sector into new status groups. Universities can then take up the banner of their niche—as a teaching focused institution or comprehensive research institution—to promote their “unique” mission through buying into rankings-related products that serve as a form of symbolic capital (Bourdieu, 1980) regardless of it having any real meaning or value beyond its role in signifying status within the niche. An unintended consequence may be that once a university has cultivated its niche as an institutional identity, it may be difficult to escape in the future.

4. Provide Clarity about the range of information sources for rankings and the messages each source generates.

This principle addresses the question of where data for rankings is coming from, and what aspect of universities each source represents. In the description for this principle, IREG uses databases, students, professors, and employers as examples of information sources. The data regarding students typically take the form of faculty to student ratio which is used as a proxy for commitment to teaching and student experience; or the percentage of international students serves as a proxy for international experience provided for the general student body or international reach of the university. These student data are typically self-reported by universities and acquired from each university's own institutional databases. While rankers typically ask for such data using a standard definition, the means of collecting, categorizing and storing data across universities is unstandardized. Universities must transform the data from the categories used in their regular operations to those requested by rankers.

There is considerable incentive to transform these data in ways that favor the university—in other words, a process that allows for “gaming” the system (e.g., Sauder and Espeland, 2009).

Data from professors and employers come in the form of university self-reports or survey data. There are reports of universities forwarding the survey requests to allies who answer the questionnaires on their behalf. Indeed, I met representatives from one Asian university at a conference, was soon on their mail list, and within a year was asked if I would be willing to respond on their behalf. Rankers can be transparent about their data sources and collection methods, but the processes that universities undertake in regard to their own data sources are usually opaque. Rankers may also recognize limitations, but such recognition leads to the same problem I articulate throughout this discussion: the ranking itself becomes all that matters. The linear order of universities is what gets attention and limitations of the data are ignored in the hierarchy of greater or less than.

5. Specify the linguistic, cultural, economic, and historical contexts of the educational systems being ranked.

This principle aims to address the complaint that rankings are biased toward English speaking and Western universities. Like many of the other principles this one runs into the problem that the ranking is a hierarchical list. People glance at the top 10, 100, or 200, then move on with their lives. Many university administrators I spoke with said that rankings are only good for knowing approximately “what league you are in”, and university promotional materials typically only cite that they are a top “x” university. Rankers could hire local historians and linguists in each region where they rank universities and create their websites so people accessing them are immersed in a state of the art learning environment that allows readers to engage with history, culture, and languages of a particular region, but this exercise is impractical. The expense would be immense, and the ranking's purpose—to function as an immediate and thoughtless assessment on the part of the reader—would be defeated.

This is also unlikely what consumers would want. One way around this complication has been to divide rankings according to cultural, linguistic, and/or socio-economic regions, such as the Times Higher Education and QS BRICS rankings or Asia rankings (THE, 2014; QS, 2014); and the US News and World Report Best Arab Region Universities (US News, 2014). However, this approach homogenizes within regions, erases history and diversity within them, and creates more niches and hierarchies.

Design and weighting of indicators

6. Be transparent regarding the methodology used for creating rankings.

Unambiguous methodology is a primary principle of scholarship and the focus of much concern in academic publishing and peer review. Not surprisingly, there has been a great deal of attention given to ranking methodologies (Hazelkorn, 2011). This principle attempts to address ongoing concerns with ranking methods. However, ranking publishers are often private-for-profit businesses and their methodologies and information sources are intellectual and commercial property that must be protected (Usher, 2014). When I asked Maclean's Magazine for an interview they were clear that they would not share information about their data or methods, because these are protected intellectual property. Rankers routinely do not report critical information such as sample size, response rates, or how missing data are handled which would allow critics to understand the representativeness and validity of results. The potential value in principle four—peer review and transparency/clarity in reporting—are not likely to be actualized since it is not in the ranker's business interests to do so.

7. Choose indicators according to their relevance and validity.

Choosing indicators based on relevance and validity simply does not happen in practice and rankers typically rely on readily available data. Data is acquired by having universities submit information that they have on hand through their own institutional databases; or rankers get data

through partnerships with publishing corporations such as Elsevier and Thomson Reuters. I have already explained the issues of non-standard categories, data collection and reporting. When evidence of the data's shortcomings or lack of validity become apparent, some limitations might be acknowledged, but they are more frequently ignored or it is argued that the data suffice.

8. Measure outcomes in preference to inputs wherever possible.

The concern with an outcomes focus is a response to arguments that focusing on inputs (e.g., student matriculation averages, student scores on standardized tests) does not reflect the character of what happens within any given institution (Marginson, 2007). An outcomes focus is believed to reflect processes that lead to particular outputs from institutions (e.g., highly desirable and adaptable employees, high quality research papers, innovative technologies). This principle seems straightforward. If we are to evaluate what happens at a particular institution, we should be concerned with what comes out of it. Yet outcomes logic remains flawed. It presumes that what happens between the point that something goes in and before it comes out is not worthy of attention. The educational and research process are black boxed so that once particular outcomes are determined and an institution can be seen to measure well on those, what occurs in the middle is presumed to not matter.

For example, US News and World Report's rankings use selectivity as an input measure. It is presumed that some processes are in place at certain universities that lead to the rejection of a greater number of students—the inputs—than other universities. They also believe that high levels of rejection indicate better quality. But if you are inputting wealthy, highly disciplined, and well-educated students, then putting out successful students leaves the question of whether your institution did much to make a difference and whether these people would have succeeded just as well at a competitor's school. The US News system rewards institutions with an elitist approach to university education and penalizes those interested in serving a wide range of students. As Gladwell (2011) asserts, "...the Yales of the

world will always succeed at the US News rankings because the US News system is designed to reward Yale-ness.” That is, a focus on particular inputs or outcomes fosters a particular model of what a university should be, if that model matches well with what certain universities—like Yale—are doing, they will rank highly by default. The rankings enact Yalness by measuring other universities against that standard. Gladwell (2011) has contrasted Yale with Pennsylvania State University (PSU), which accepts many underprivileged young people, finding that PSU has many supports in place to ensure their students succeed. Despite expectations based on the “poor quality” input at PSU, they have outstanding graduate outcomes. The US News rankings do not capture this relationship between input, process, and outcome.

9. Make the weights assigned to different indicators (if used) prominent and limit changes to them.

This principle responds to complaints around methodological clarity and recognition that weightings determine ranking results. Making weights prominent and limiting changes is intended to provide clarity on what indicators are used, what they mean, and how they advance a particular vision of the university. The “Rankings Game” website (Stake, 2015) allows visitors to create their own law school rankings by allocating different weights to various measures. Gladwell (2011) has generated a top 10 law schools list by equally weighting tuition price, academic reputation, LSAT (law school admissions test) scores at the 75th percentile, student-faculty ratio, and faculty law-review publishing. By giving a forty percent weight to price, forty percent to LSAT scores, twenty percent to publishing, and removing the other measures the top 10 list changed dramatically. The University of Alabama suddenly appeared in the list sandwiched between the University of Colorado (ranked 7) and Stanford University (ranked 9).

How particular measures are weighted will determine ranking results and the reason changes to weights should be limited is to ensure comparability on a year-to-year basis. This principle, as a general

assertion and practice is agreeable, but again, most ranking consumers will not be well-trained in research methods. Moreover, because rankings are a technology of simplification, bogging a reader down in details about weightings will defeat the purpose of knowing at a glance and works against their real world application.

Collection and processing data

10. Pay due attention to ethical standards and the good practice recommendations articulated in these Principles.

This principle appears to articulate a distinction between “ethical standards” and “the good practice recommendations” of the Berlin Principles. Yet outside of the Berlin Principles, there are no clear ethical standards for ranking and IREG does not provide any direction. The Berlin Principles are the only ethical standards associated with university rankings. Curiously, the description for this principle does not speak to ethics, but adds more to ranking methods. It states that, “In order to ensure credibility...those responsible for collecting and using data...should be as objective and impartial as possible.” Objectivity can be an ethical standard, but how it is meant in this instance is unclear.

Sayer (2000) identifies at least three distinct meanings for the concept of objectivity: 1) value neutral, indifferent or value-free; 2) searching for true or practically adequate knowledge; 3) referring to the nature of objects independent of what any person may think of them (p. 58). These interpretations of objectivity are often confused or conflated. For example, although it is presumed that we must first have value neutral inquiry to achieve adequate knowledge (objective 2), this assumption is false. We can find practically adequate knowledge from value-based inquiry (Sayer, 2000). The third form of objectivity requires a question of whether we can know something independent from our thoughts about it—and of course we cannot—but we can distinguish between more or less adequate statements regarding the object of concern.

The relationship between university rankings and objectivity is complicated. To start, universities and rankings are constructed conceptually and physically through much heterogeneous engineering (Law and Hassard, 1999). The process begins with people developing a concept of a university and then seeking evidence that the university exists according to measures for that concept. Such inquiry and the transformations it involves are value-laden, not value-neutral. In other words, rankings cannot be objective in the sense of being value-neutral. However, university rankings can achieve objectivity in the sense that they can be practically adequate knowledge referring to phenomena of interest. This means that what values we use to evaluate, audit, and organize universities will have particular effects on how it is enacted and particular consequences for society more broadly.

11. Use audited and verifiable data whenever possible.

A cynical interpretation of this principle is that it exists merely to serve IREG's purposes of selling an audit service. After all, auditing “has always been a business” (Power, 2003:382). The service is provided to IREG members for 4000 Euro, or 6000 for non-members (IREG, 2014). Such a practice not only ensures IREG's ongoing relevance, but also provides another form of symbolic capital in that successful audit allows rankers to wear an IREG badge of approval. The audit and the badge symbolize and perform transparency while also signifying legitimacy. However, the Berlin Principles themselves do not refer to IREG's auditing process, merely to the advantages of audited data in that will be more agreeable to institutions to which the data refer, and allow comparability.

To suggest that one has data agreeable to all ranked institutions, and that such are comparable is a tremendous task. It would require that all ranked institutions are consulted about the categories of data used in the rankings, and that each institution operates according to such categories. The first requirement addresses agreeableness, and the second comparability. In reality, what typically happens is that rankers have pre-established categories that they require institutions to submit to, whether or not

they agree. Institutions are left to figure out how to make themselves fit into the categories. Moreover, to my knowledge the Times Higher Education is the only ranking organization that will not rank a university if it asks to be removed from their system. All other ranking organizations will seek out data and rank institutions if they refuse to submit. Early in my interviews with deans, department chairs, and institutional analysis staff, I learned a widespread problem is that comparable categories are rare. Universities have different means of categorizing students, professors, research funding, and organizational units. The work of submitting to the categories provided by rankers produces a standardized and comparable form, but that process produces the appearance of comparability in instances where institutions may not be comparable at all. This is one process through which rankings homogenize, simplify, and enact university identity.

12. Include data that are collected with proper procedures for scientific data collection.

Principle twelve is another aimed toward addressing method critiques. IREG explains that “using data from unrepresentative or skewed subsets of students, faculty, or other parties may not accurately represent an institution or program and should be excluded.” I have dealt with this principle to some degree in my discussion of the others above. Suffice to say that they do not report on how many of each category they consult with, nor do they include their response rates. From whom they acquire their data and the representativeness of such groups is routinely made opaque.

13. Apply measures of quality assurance to ranking processes themselves.

The descriptor for this principle refers to ensuring expertise of the people doing ranking, and ongoing consultation with experts to evaluate the rankings. That is, rankings should follow good methodological principles, undergo audit, and be improved upon. Again, the cynical interpretation of

this would be that this principle serves to have organizations submit to IREG's audit process or at least consult with its members, since they are the International Ranking Expert Group. If a ranking organization follows all of the Berlin Principles and is able to do so in ways that address their limitations, they would be doing much quality assurance. The most effective means of doing so would be for rankers to engage with the institutions they rank and adapt their methods accordingly. Though such a practice contradicts the Berlin Principles that strive for comparability.

14. Apply organizational measures that enhance the credibility of rankings.

This principle relates closely to principle 13 in that it articulates a need for consultation with advisory bodies, “preferably with some international participation” (IREG, 2014). To be clear, the majority of ranking organizations are private corporations engaging audiences as they wish. Some choose to engage directly with universities and make adjustments. However, this positions scholars and university administrators as stakeholders and advisors where they may have had no interest in advising or holding stake in the consequences that rankings create. Another problem with this principle is that enhancing credibility may have little to do with enhancing ethics, validity, accountability or other matters of concern that come to mind when one thinks of credibility. Instead, enhancing credibility may involve producing symbolic performances that appear to address issues of concern with little in the way of action to resolve them—a form of decoupling. Indeed, Power (2003) has demonstrated that closure of controversy and image management to enhance credibility are primary to legitimization. The Berlin Principles gesture toward ethics and perform a professional identity, but are decoupled from practice.

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.

Ranking displays merely allow users to create alternative rankings. The one ranking system

that attempts to innovate its displays is U-Multirank, which allows users to select criteria most important to them, then illustrates a table with a letter grade for each criterion. The user then judges which institution is best (U-Multirank, 2014). U-Multirank allows variety in how rankings are displayed, and does not provide the option for a single linear comprehensive ranking in the typical format. The net result is a hierarchy created through a combination of pre-determined categories, user values, and judgment. The novelty of this approach is that it makes the user's value system visible along with characteristics that make-up a university. Yet realizing these unique qualities requires that users consider all of the categories and not immediately select a few for their assessment. Engaging with U-Multirank becomes a learning experience, rather than a passing judgment. This defeats the efficiency and clarity of rankings in that the user not only has to think heavily on matters of complexity and importance, but also to practice using the tool and to learn how to interpret the results. U-Multirank is exemplary of addressing the fifteenth principle, but is costly, time intensive, and defeats the efficiency that give rankings their purpose.

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.

A primary problem with data is a lack of standardization across universities in terms of how they categorize and define their operations. The other problem with this principle is that once ranking results have been published, consequences will have had their effects. Rankings have come to carry such weight in some circles that careers can be lost (Hazelkorn, 2011), funding for particular programs can be cut, and student applications for scholarships or student loans to attend particular institutions may be denied. If rankings are readjusted after these consequences, such adjustments are unlikely to restore things to what they once were.

Summary and Conclusion

At the most basic level, the Berlin principles and its auditing system are technologies of trust that simultaneously provide audiences with a sense that ranking processes are transparent, ethical, and effectively policed (Porter, 1995). However, it should now be clear that the Berlin Principles typically contradict the logic and practice of ranking. The practical purpose of rankings is to create rapid understanding without reading into detail and transcend local particularities in order to avoid complications inherent in each. If applied in practice, the Berlin Principles would complicate a technology of simplification. While they may be used to audit rankings to some degree, they largely cannot be effectively realized because the technology they are intended to police, the infrastructure upon which that technology rests, and the ways people use the technology are not amenable to such principles. Rankings cannot be tightly coupled with the Berlin Principles.

As Power (2003) has articulated, audit involves a great deal of image management, and it is an expressive process. The Berlin Principles are an industry's symbolic response to intensified criticism and are a gesture toward scientific standards that tempt to enact a scholarly and scientific identity. The Berlin Principles lend legitimacy to rankings without delivering on the practices they promise. Given that the Berlin Principles are decoupled from actual ranking practice we can only know ranking organization's true ethical principles by observing their actions. Since their actions do not match with their principles, these cannot be their true principles (Barker, 2008). Their actions are dissonant with the values and evaluation processes of traditional academia. Rather, the Berlin Principles signify legitimacy and commitment to transparency by defining an ideal role for IREG's audit and rankings more broadly, thereby enacting the institutionalization of rankings business and associated practices.

The Berlin Principles are also a response to representational risk inherent in trying to evaluate academic communities whose scientific practices—such as in tenure, promotion, and grant/publication peer review (Lamont, 2010)—are dissonant with rankings business. Academic culture promotes “the

best” based on diverse value systems thereby allowing diversity to grow. This “structural accretion” is the way in which universities are able to multiply organizational missions and functions while maintaining traditional interests (Smelser, 2013). Rankings are a technology that erase diversity, promote homogeneity, and obfuscate methodology, the very issues the Berlin Principles aim to address.

Many of the principles also serve the interests of ranking organizations—sometimes also those of their audiences—in that they contribute to the proliferation of rankings, related products, and industries. Rankings and associated practices create new problems and questions for universities as they simultaneously offer solutions. For example, each ranking will choose a specific set of outcomes to measure based on their model of what a university should be. Any set of outcomes will be isomorphic with the model by which a university operates to a greater or lesser degree. The degree of mismatch between the two will determine high or low rank. The degree to which a university strives to rank well—due to internal decisions or environmental factors such as government policy, aims to cultivate an identity, or concern for reputational risk—will lead to increasing isomorphism. Where there is persistent anti-isomorphism, there will be opportunities for new ranking products as new niches are identified. This is not merely a matter of rankings creating coercive isomorphism (Dimaggio and Powell, 1983) that incentivize their model of evaluation. Rather, some ranking values are also frequently held by higher education institutions and this makes them a potentially legitimate goal. Academic criticisms of rankings are largely due to dissonance between their respective value systems, but such dissonance can be rationalized by an appeal to legitimate alternative values. What I have described are processes that clarify the homogenizing effects that rankings have on universities, the proliferation of rankings, and related business.

While rankings hold values that may be well aligned with some in traditional scholarly communities, they fundamentally transform how valuation occurs and thereby change what is possible to do in higher education institutions. Instead of structural accretion and diversity, rankings promote

homogeneity and exclusion by narrowing what can be considered as valuable, imposing narrow categories of thought, promoting a form of competition that is unusual to academic institutions, and positioning academics as stake holders in games to be played by purchasing analytic products that further rank style evaluation and thinking.

The processes I have discussed here transform the field of higher education into a linear hierarchy and constitute a complex cultural economy composed of many heterogeneous parts (Du Gay and Pryke, 2002) that incorporate academic practices, but transforms them into something different. The symbolic components of this are important as the identities of ranking organizations become defined through texts such as the Berlin Principles and those of countries, universities, and their constituents become defined by rankings, shaping higher education along the way. Rankings will continue to have force because of the industries and wealth they generate, but also because they adopt values and a discourse of assessment that has been a part of university culture for centuries. In considering what sorts of higher education institutions we want, what contributions to society and the economy these will make, we must consider how we go about valuing and evaluating such institutions.

References

- Amsler, S.S., & Bolsmann, C. 2012. "University Ranking as Social Exclusion." *British Journal of Sociology of Education* 33(2):283–301.
- Barker, D. 2008. "Ethics and Lobbying: The Case of Real Estate Brokerage." *Journal of Business Ethics* 80(1):23–35.
- Baty, P. 2014. "The Times Higher Education World University Rankings, 2004-2012." *Ethics in Science and Environmental Politics* 13(2):125–30.
- Bourdieu, P. 1980. *The Logic of Practice*. Stanford, CA: Stanford University Press.
- Bowker, G.C. & Star, S.L. 1999. *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: The MIT Press.

DiMaggio, P.J., & Powell, W.W. 1983. "The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields." *American Sociological Review* (2):147.

Du Gay, P., & Pryke, M. 2002. *Cultural Economy: Cultural Analysis and Commercial Life*. Thousand Oaks, CA: Sage.

Gladwell, M. 2011. "The Order of Things." *The New Yorker*, February 14. Retrieved January 16, 2014 (http://www.newyorker.com/reporting/2011/02/14/110214fa_fact_gladwell).

Hazelkorn, E. 2011. *Rankings and the Reshaping of Higher Education: The Battle for World-Class Excellence*. Houndmills, Basingstoke, Hampshire; New York: Palgrave Macmillan.

International Ranking Expert Group. 2014. "Berlin Principles on Ranking." Retrieved January 29, 2014 (http://www.ireg-observatory.org/index.php?option=com_content&task=view&id=41&Itemid=48).

International Ranking Expert Group. 2015. "IREG Approved." Retrieved March 24, 2015 (<http://www.ireg-observatory.org/en/index.php/information>).

Lamont, M. 2010. *How Professors Think : Inside the Curious World of Academic Judgment*. Cambridge, Mass. ; London : Harvard University Press, 2010.

Law, J., & Hassard, J. eds. 1999. *Actor Network Theory and After*. Oxford, UK: Blackwell Publishing.

Liu, N.C., & Cheng, Y. 2005. "The Academic Ranking of World Universities." *Higher Education in Europe* 30(2):127–36.

Lynch, M. 1993. *Scientific Practice and Ordinary Action: Ethnomethodology and Social Studies of Science*. Cambridge, UK: Cambridge University Press.

Marginson, S. 2007. "Global University Rankings: Implications in General and for Australia." *Journal of Higher Education Policy and Management* 29(2):131–42.

Marope, M.P. T., Wells, P.J., & Hazelkorn, E. 2013. *Rankings and Accountability in Higher Education: Uses and Misuses*. United Nations Educational Scientific and Cultural Organization. Retrieved December 19, 2014 (<http://www.unesco.org/new/en/education/resources/in-focus-articles/rankings/>).

Nigavekar, A. 2012. "Obsession with university rankings". *Financial Chronicle*, 3 April.

Power, M.K. 2003. "Auditing and the Production of Legitimacy." *Accounting, Organizations and Society* 28(4):379–94.

Power, M., Scheytt, T., Soin, K., & Sahlin, K. 2009. "Reputational Risk as a Logic of Organizing in Late Modernity." *Organization Studies* 30(2-3):301–24.

Quacuarelli Symonds. 2014. "QS University Rankings: BRICS 2014 | Top Universities." Retrieved December 2, 2014 (<http://www.topuniversities.com/university-rankings/brics->

rankings/2014#sorting=rank+country=+stars=false+search=)

Sauder, M., & Espeland, W.N. 2009. "The Discipline of Rankings: Tight Coupling and Organizational Change." *American Sociological Review* 74(1):63–82.

Sayer, A. 2000. *Realism and Social Science*. Thousand Oaks, CA: SAGE Publications.

Sayer, D. (2016). *Rank Hypocrisies: The Insult of the REF*. London, UK: SAGE Publications. Retrieved from <https://us.sagepub.com/en-us/nam/rank-hypocrisies/book244105>

Smelser, N.J. 2013. *Dynamics of the Contemporary University: Growth, Accretion, and Conflict*. University of California Press. Retrieved June 14, 2015 (<http://california.universitypressscholarship.com/view/10.1525/california/9780520275812.001.0001/upso-9780520275812>).

Smith, D.E. 2006. *Institutional Ethnography as Practice*. Lanham, Md.: Rowman & Littlefield.

Strathern, M. 1997. "Improving Ratings': Audit in the British University System." *European Review* 5(3):305–21.

Stake, J. 2015. "Play The Ranking Game: The Law School Ranking Game Page: Indiana Law." Retrieved March 28, 2015 (<http://monoborg.law.indiana.edu/LawRank/play.shtml>).

Stark, D. 2009. *The Sense of Dissonance. [electronic Resource] : Accounts of Worth in Economic Life*. Princeton, NJ:Princeton University Press.

Times Higher Education. 2014. "World University Rankings - Home - Times Higher Education." Retrieved December 16, 2014 (<http://www.timeshighereducation.co.uk/world-university-rankings/>).

Usher, A. 2014. "The Problem with Global Reputation Rankings | HESA." Retrieved October 9, 2014 (<http://higheredstrategy.com/the-problem-with-global-reputation-rankings/>).

US News. 2014. "Top Arab Region Universities | US News Best Arab Region Universities - US News." *US News & World Report*. Retrieved December 16, 2014 (<http://www.usnews.com/education/arab-region-universities>).

Walker, Sally. 2004. "Diversity Not Being Considered". *The Australian*, December 1, 42.

West, P.W.A. 2009. "A Faustian Bargain? Institutional Responses to National and International Rankings." *Higher Education Management and Policy* 21(1):9–18.