August 2017. Prepared for an ASA Author Meets Critics session on Wendy Espeland and Michael Sauder's *Engines of Anxiety: Academic Rankings, Reputation, and Accountability* (Russell Sage, 2016).

By the Numbers

Kieran Healy
Duke University
kjhealy@soc.duke.edu

The end of June meant exam results. I was an undergraduate at University College Cork, in Ireland. Everyone sat their examinations at the beginning of the month. The scripts were blind-graded by two markers, discussed by a committee with an external examiner, and then posted. Eventually a letter would arrive at your house with your scores. But first, and much more quickly, they went up in the corridor of the main quad. Large sheets of tractor-fed printer paper were pinned behind glass on the official announcement boards, grouped by faculty—Arts, Science, Law, Medicine, and so on. They showed every student's grades for each of their degree subjects, from F for fail and P for pass on up to 1H for First Class Honors. Although students had ten-digit ID numbers, the reporting system helpfully printed out everyone's name in full. You went over to campus to look yourself up, look up the names of your friends, and perhaps also those of your enemies. There were two lists. The first was sorted alphabetically. The second was ranked, running from the student with the highest overall score down to the lowest honors mark.

When you have been educated in a system that reports its results in this way, you are very aware of the power of quantitative assessments of quality, of the ease with which public rankings can become widely known, and of the capacity for league tables to produce and sustain an unpleasantly febrile state of mind amongst the people being evaluated. Thus, by training and disposition I am very sympathetic to the argument made by Wendy Espeland and Michael Sauder in *Engines of Anxiety*. It is an excellent

book that will stand for some time as a definitive assessment of the fortunes of U.S. law schools in the era of rankings, and also as a significant contribution to a literature on commensuration and valuation that one of the authors herself pioneered. I generally believe what it has to say about the effect that *U.S. News and World Report*'s rankings (henceforth *USN*) had on law schools, effects ranging from additional pressures on administrators and faculty, to forced and perhaps wasteful changes in budgets, to changes in the choices facing applicants and the chances facing graduates. I am also in broad agreement with the way Espeland and Sauder think about rankings on the whole, and in particular with their view of the strange, faintly absurd, but seemingly inescapable trap that *USN* set for law schools. The academic legal establishment did not so much fall into this trap as become entangled in it. Like a fly touched by the thread of a spider's web, they were at first only lightly caught up, but then found that each move they made in response only drew them in more tightly.

The heart of the book is a puzzle about the loss of a profession's control over its ability to make authoritative judgments about quality and prestige in its own domain. Whether seen as a rational investment in one's human capital, the driveshaft of credentialed reproduction and social inequality, or the institutional expression of a cultural myth of abstract universal knowledge, social theory has long placed universities at the core of modern societies. At the core of universities are experts who have successfully claimed the right to judge what counts as knowledge, who have defined the scope and power of abstract ways of knowing, and established the criteria for professional authority, often backed up through state-sponsored credentials and certification. And yet, as noted at the beginning of the book (quoting Bard College's president, Leon Botstein), whole swathes of higher education seems to have been "bamboozled by a third-rate news magazine" (1). How could this happen?

The authors approach the problem constituency by constituency. There is a chapter on how prospective law students use the rankings; one on how rankings affect the admissions process and the staff in admissions offices; then a chapter on how rankings change the job of Deans both with respect to the people that report to them and the university authorities, donors, and alumni they must face; and a chapter on the effect of rankings on job placement and the career prospects of law graduates. These detailed discussions—relying on a comprehensive, long-term program of in-depth interviews and analysis of primary textual material—are bookended at the beginning by a general analysis of the nature of numerical rankings, and at the end by a discussion of their presence and prospects in other areas of life.

The opening analysis is set up in terms of the demand for "accountability" from various interested parties (or "stakeholders", as they are often called), coupled with an argument about the distinctive character of quantitative or numerical assessments in making people or organizations accountable. Espeland and Sauder argue that numbers seem objective and authoritative, that they simplify and organize otherwise opaque

fields for the purposes of outside assessment and control, and that they provide a basis for decisions that might otherwise rely on guesswork or unfounded authority. The wrinkle is that these measures feed back on the system they assess. In the authors' terms they are *reactive*: they change what they measure. This happens in several ways. First, because they only measure certain features of a law school, and those only in specific ways, rankings encourage schools to value, focus on, and shore up whatever produces the measure. Things that are not counted will not count. Second, measurement flattens out differences in character into differences of magnitude. Everything is fed into a score. All that matters is how large a contribution it makes to the final result. Third, and even worse than focusing only on what is measured, the ranking itself may become a policy target, as opposed to a byproduct of running a good law school. This results in various perverse or self-fulfilling outcomes, as administrators spend time and money attempting to game key measures (and keep their consciences clear). Finally, as a result of all this effort, the rankings become even more salient to all those affected by them. They tend to become the main thing that schools tell stories to themselves and others about, acting both as a coordinating device for action and a powerful, focal symbol that gives organizations and other actors their sense of strategy and reason.

In their closing discussion, Espeland and Sauder survey the rise of rankings and scoring across many different fields ranging from undergraduate education in the U.S. to crime statistics for large cities. They note the differences across these different settings, mostly in the precise ranking method used, and sometimes even in the successful rejection of external ranking altogether. Dental schools managed to collectively organize a resistance to being ranked by *USN* at all. But that sort of result is rare. More often, rankings—whether officially mandated, generated by magazines, crowdsourced, or self-administered—win out in some form. This "quantitative accountability" is on the rise everywhere. While acknowledging that rankings can be useful, Espeland and Sauder note that "useful is not the same as good" (198), and they worry about the morally corrosive effects of rankings. These methods "tend to foreclose other modes of creating and expressing value", in particular "those forms of evaluation that are messier and more time consuming and that require deliberation or consensus" (201). They also make for cynicism in administration and erode the responsibility, discretion, and ability to exercise one's independent judgment that is a hallmark of professional authority. "[A]ccountability measures must themselves be held accountable and made transparent" they suggest, and there shoudl be "thoughtful consideration about the ethical aspects of accountability measures, so that quantitative evaluations are not assessed solely on the uses they serve" (201).

The degree to which the moral critique of ranking methods overlaps with the self-interest of administrators is a delicate point, and the book finesses it a little. The authors visit each constituency in turn, from applicants to Deans, and carefully demonstrate the occasionally useful but mostly pernicious effects of the *USN* rankings. Similarly,

they find that people in these different locations are occasionally supportive but mostly annoyed or even despairing of the ranking system. The case they make can be thought of as having three components. The first asks, "What do rankings do?" This question is about how rankings affect administrators, how they are used and interpreted by everyone involved, how they act as a point of coordination across a field, and how, once in motion, attention to them can restructure individual choices and institutional budgets, especially through various kinds of perverse incentives to manage or game the system. The book makes its case very convincingly here. The interview material is excellent and the authors make expert use of it. The strategy of surveying each constituency in turn results in a comprehensive survey of just how much rankings changed things right across the field of legal education.

The second component of the argument asks, "Why does this happen?" Why are these rankings so powerful? I found some parts of the case here more plausible—or at least easier to see—than others. The relationships between reactivity in general, quantification in general, and ranking in particular are not always so clear. The opening discussion focuses on how numbers and measurement in general are good for producing visibility and accountability, especially for outsiders. In addition, Espeland and Sauder note that "making numbers requires categories with strong boundaries" (23) and that their cultural authority can "bolster our courage to act" (21). Despite my basic sympathy for the argument, I found myself asking whether numbers are really that much more powerful than alternative methods of coordination and control, especially when it comes to generating the reactivity that the argument emphasizes. Catechisms, shibboleths, purity regimes, ritual compliance and symbolically-infused hierarchy in all its forms can ease decision-making, clarify boundaries, and produce portable tests of success, and displace action from meaningful substance to hollow performance. All of them are reactive, in the sense that—like many social processes—they simply require some mutual orientation towards them in order to have the potential to change the minds, plans, and trajectories of those involved. Insofar as they are about distinguishing better from worse performance, as opposed to simply affirming the uniqueness of every single thing in the world, any qualitative mode of classification could be just as powerfully commensurating and just as prone to reactive, self-fulfilling mechanisms. At times, the emphasis on reactivity seems less like an explanation and more like a restatement of the outcome.

One reason for this is the occasional shift in the general argument from the nature of rankings in particular to the character of counting, measurement, or quantification in general. The general effects of quantification seem more incohate, varied and inconsistent than the particular power of rankings. However, the book does put its finger on one of the distinctive features of organizational responses to ranking systems. We know from decades of research by institutionalists that, whether qualitatively or quantitatively assessed, organizations are happy to generate and comply with all

manner of reporting requirements while also decoupling their actual activities from such formalities. For example, in the process of renewing its accreditation a university might generate a vast quantity of data, reports, and plans that are gathered up by the administration and presented to the relevant authorities. Tedious, to be sure, but also in general securely insulated from much of what happens inside of classrooms. A public ranking that people take seriously is not like this. It cannot be shelved or ignored in the same way. It provides a link between experts and their audiences that is both coercive and seductive. It does this by forcing the school into a seemingly clear relationship with its peers, peers that can now much more easily seen by insiders and outsiders alike as competitors. It is here that there does seem to be something distinctive about rankings in particular, something that goes beyond the collection of data, a focus on numbers, or a love of quantification. Rankings can bypass conventional methods of merely symbolic compliance, because one cannot merely comply—one is also situated on a scale.

Even here, some prerequisites must be met before this can happen. Symbolic compliance and decoupling can reassert themselves. The more rankings there are, for example, the less disciplining any one of them can be. If there are too many competing measures of overall quality, the entire business of ranking may lose its legitimacy. If there are too many measures, organizations can defuse the problem of competition, and slip back towards a comfortable system of compatible niches, by choosing to highlight the measure of performance it likes best. But if there is only one ranking that matters, then it may well have its distinctive effects by ordering every unit in terms of the others. That is quite different from a threshold that any organization can meet, or an award that can be won by a single organization without impugning the respectability of all the other players. The authors make this point early on and it seems to me to be one of the book's key contributions, a little-recognized but vital element that really helps us understand why rankings can take hold as they do.

The distinction between merely symbolic organizational compliance (often with some general set of criteria or requirements, perhaps numerically expressed) and the more severe world of direct competition over rank position brings us to the third component of the book's argument. This is the moral element, the question of whether rankings are a useful tool or a severe threat to important modes of valuation, as well as a danger to the independent ability of experts to exercise their own judgment in this regard. Again, the argument leans a little too much toward a general critique of quantitative comparison or valuation when perhaps more might be gotten from a consideration of rankings in particular. Rankings introduce a form of compliance that professionals resist and find morally offensive. Professional authority is rooted in the independent exercise of good judgment by credentialed experts. The credential is grounded in a university or professional society, and usually backed up by the state, but it in general membership is not graded or ranked. As the old saw goes, "What do you

call the person who graduated last in his class in medical school?" Answer: "Doctor." Meeting the threshold is all that matters.

Now, it is tempting to think that what is pernicious about rankings is something that's also true about numbers in general, as the book sometimes suggests. However, professionals typically try to reject *any* control device that impinges on their discretionary judgment, whether it takes a numerical form or not. Mandatory check-lists in airline cockpits, or more recently in operating theaters, are an example of the sort of thing that provokes fierce resistance. Moreover, it is not even that professionals dislike making judgments about the quality of their peers. Informal status competition and ranking is rife in professional and especially academic circles. But while it is partly formalized in the world of prizes and awards, systematic efforts to score and rank peers tend to be refused. It seems indecent. Everyone has a private assessment of quality, but no measure is thought sophisticated enough to capture it.

The dilemma, or the moment of bad faith when it comes to the moral argument, is that professionals cannot be professional except through the exercise of their judgment; their judgment is very often expressed through an implicit or explicit ranking; and yet they do not want to be subject to any such formal ranking themselves. Faculty, deans, and administrations spend most of their time sorting and ranking students. Internally, they administer tests and hand out grades. Externally, they rely very heavily on standardized test scores for admission to undergraduate, graduate, and professional programs of study. When alternative modes of valuation are employed in these settings, as when some schools stop awarding grades, it is generally seen as an annoyance by everyone else. When the desire to sharply judge and precisely rank students seems to flag, as when 75 percent of students average an A-minus grade, then it's generally agreed that something has gone wrong with the system.

Everyone—including the faculty at law schools—wants a flow of actionable information about the targets of their decision-making. They want some clear signal of what they are dealing with in order to exercise their own judgment. If the information supplied is quantitative, so much the better. But it need not be. Maybe what matters is where someone went to school, or who their advisor is, or what their religion is, or whether they were in the right secret society, or whether they have been ritually purified by the correct rites. Few people want to be assessed on a reductive basis, at least not very often. That makes it tempting to argue against the idea of ranking in general. But then the moment of bad faith returns, because academics and other professionals really do believe that some of the work put out by their peers is better than the alternatives. To remain a professional at all they must believe in their own capacity for judgment on such matters. Status in these settings condenses out of the high regard of highly-regarded others, and it produces implicit, loose, but real rankings all the time. That makes a *general* moral objection to rankings harder to sustain in a full-throated way.

The end of June meant exam results. The tension that came with the walk down the corridor was palpable, especially for those close to any of the boundaries, whether you were hoping to scrape a pass or desperate to find your name toward the top of the list. It was unfair; it was prone to error; it was invidious and even personally destructive. It was not what the life of the mind was meant to be about. But you still cared, and you still believed, because you believed that your work was better, that your ideas were better, than at least some of what was on offer. Even if you felt that the system was wholly unable to assess the quality of your ideas—that it had failed to grasp your singular genius, that it had missed the groundbreaking character of your research, that it was chronically prejudiced against people like you—what you wanted was not an end to the idea of quality, but a proper reformulation of it, a true appreciation of your contribution relative to that of others. Law school faculty, deans, and administrators are long past the time of their lives when their individual performance is routinely assessed in terms of As or Bs, as *magna* or *summa cum laude*. But they still exercise that capacity for judgment over others every week of the semester. They believe in it. They are still committed to the view that they know and can assess quality when they see it, and they usually think they can reliably quantify it. It's just that they would rather not be subject to that pressure themselves. Becoming a faculty member should have been a way to escape it. The injustice, or at least the indignity, at the heart of a ranking system is the experience of having one's own knife turned back upon oneself, and finding that it still cuts like it used to.