Car and Driver conducted a comparison test of three sports cars, the Lotus Evora, the Chevrolet Corvette Grand Sport, and the Porsche Cayman S. Yet when you inspect the magazine’s tabulations it is hard to figure out why Car and Driver was so sure that the Cayman is better than the Corvette and the Evora. The trouble starts with the fact that the ranking methodology Car and Driver used was essentially the same one it uses for all the vehicles it tests—from S.U.V.s to economy sedans. It’s not set up for sports cars. Exterior styling, for example, counts for four per cent of the total score. Has anyone buying a sports car ever placed so little value on how it looks? Similarly, the categories of “fun to drive” and “chassis”—which cover the subjective experience of driving the car—count for only eighty-five points out of the total of two hundred and thirty-five. That may make sense for S.U.V. buyers. But, for people interested in Porsches and Corvettes and Lotuses, the subjective experience of driving is surely what matters most. In other words, in trying to come up with a ranking that is heterogeneous—a methodology that is broad enough to cover all vehicles—Car and Driver ended up with a system that is absurdly ill-suited to some vehicles.

A heterogeneous ranking system works if it focuses just on, say, how much fun a car is to drive, or how good-looking it is, or how beautifully it handles. The magazine’s ambition to create a comprehensive ranking system—one that considered cars along twenty-one variables, each weighted according to a secret sauce cooked up by the editors—would also be fine, as long as the cars being compared were truly similar. It’s only when one car is thirteen thousand dollars more than another that juggling twenty-one variables starts to break down, because you’re faced with the impossible task of deciding how much a difference of that degree ought to matter. A ranking can be heterogeneous, in other words, as long as it doesn’t try to be too comprehensive. And it can be comprehensive as long as it doesn’t try to measure things that are heterogeneous. But it’s an act of real audacity when a ranking system tries to be comprehensive and heterogeneous—which is the first thing to keep in mind in any consideration of U.S. News & World Report’s annual “Best Colleges” guide.

The *U.S News* rankings… relies on seven weighted variables:

1. Undergraduate academic reputation, 22.5 per cent

2. Graduation and freshman retention rates, 20 per cent

3. Faculty resources, 20 per cent

4. Student selectivity, 15 per cent

5. Financial resources, 10 per cent

6. Graduation rate performance, 7.5 per cent

7. Alumni giving, 5 per cent

From these variables, *U.S. News* generates a score for each institution on a scale of 1 to 100,

This ranking system looks a great deal like the Car and Driver methodology. It is heterogeneous. It doesn’t just compare U.C. Irvine, the University of Washington, the University of Texas-Austin, the University of Wisconsin-Madison, Penn State, and the University of Illinois, Urbana-Champaign—all public institutions of roughly the same size. It aims to compare Penn State—a very large, public, land-grant university with a low tuition and an economically diverse student body, set in a rural valley in central Pennsylvania and famous for its football team—with Yeshiva University, a small, expensive, private Jewish university whose undergraduate program is set on two campuses in Manhattan (one in midtown, for the women, and one far uptown, for the men) and is definitely not famous for its football team.

The system is also comprehensive. It doesn’t simply compare schools along one dimension—the test scores of incoming freshmen, say, or academic reputation. An algorithm takes a slate of statistics on each college and transforms them into a single score: it tells us that Penn State is a better school than Yeshiva by one point. It is easy to see why the U.S. News rankings are so popular. A single score allows us to judge between entities (like Yeshiva and Penn State) that otherwise would be impossible to compare.

A comprehensive, heterogeneous ranking system was a stretch forCar and Driver—and all it did was rank inanimate objects operated by a single person. The Penn State campus at University Park is a complex institution with dozens of schools and departments, four thousand faculty members, and forty-five thousand students. How on earth does anyone propose to assign a number to something like that?

The first difficulty with rankings is that it can be surprisingly hard to measure the variable you want to rank—even in cases where that variable seems perfectly objective. There’s no direct way to measure the quality of an institution—how well a college manages to inform, inspire, and challenge its students. So the U.S. News algorithm relies instead on proxies for quality—and the proxies for educational quality turn out to be flimsy at best.

Take the category of “faculty resources,” which counts for twenty per cent of an institution’s score. “Research shows that the more satisfied students are about their contact with professors,” the College Guide’s explanation of the category begins, “the more they will learn and the more likely it is they will graduate.” That’s true. According to educational researchers, arguably the most important variable in a successful college education is a vague but crucial concept called student “engagement”—that is, the extent to which students immerse themselves in the intellectual and social life of their college—and a major component of engagement is the quality of a student’s contacts with faculty. As with suicide, the disagreement isn’t about what we want to measure. So what proxies does U.S. News use to measure this elusive dimension of engagement? The explanation goes on:

“We use six factors from the 2009-10 academic year to assess a school’s commitment to instruction. Class size has two components, the proportion of classes with fewer than 20 students (30 percent of the faculty resources score) and the proportion with 50 or more students (10 percent of the score). Faculty salary (35 percent) is the average faculty pay, plus benefits, during the 2008-09 and 2009-10 academic years, adjusted for regional differences in the cost of living. . . . We also weigh the proportion of professors with the highest degree in their fields (15 percent), the student-faculty ratio (5 percent), and the proportion of faculty who are full time (5 percent).”

This is a puzzling list. Do professors who get paid more money really take their teaching roles more seriously? And why does it matter whether a professor has the highest degree in his or her field? Salaries and degree attainment are known to be predictors of research productivity. But studies show that being oriented toward research has very little to do with being good at teaching. Almost none of the U.S. News variables, in fact, seem to be particularly effective proxies for engagement. As the educational researchers Patrick Terenzini and Ernest Pascarella concluded after analyzing twenty-six hundred reports on the effects of college on students:

“After taking into account the characteristics, abilities, and backgrounds students bring with them to college, we found that how much students grow or change has only inconsistent and, perhaps in a practical sense, trivial relationships with such traditional measures of institutional “quality” as educational expenditures per student, student/faculty ratios, faculty salaries, percentage of faculty with the highest degree in their field, faculty research productivity, size of the library, [or] admissions selectivity.”

There’s something missing from that list of variables, of course: it doesn’t include price. That is one of the most distinctive features of the U.S. News methodology. Both its college rankings and its law-school rankings reward schools for devoting lots of financial resources to educating their students, but not for being affordable. Why? Morse admitted that there was no formal reason for that position. It was just a feeling. “We’re not saying that we’re measuring educational outcomes,” he explained. “We’re not saying we’re social scientists, or we’re subjecting our rankings to some peer-review process. We’re just saying we’ve made this judgment. We’re saying we’ve interviewed a lot of experts, we’ve developed these academic indicators, and we think these measures measure quality schools.”

As answers go, that’s up there with the parental “Because I said so.” But Morse is simply being honest. If we don’t understand what the right proxies for college quality are, let alone how to represent those proxies in a comprehensive, heterogeneous grading system, then our rankings are inherently arbitrary. All Morse was saying was that, on the question of price, he comes down on the Car and Driver side of things, not on the Consumer Reports side. U.S. News thinks that schools that spend a lot of money on their students are nicer than those that don’t, and that this niceness ought to be factored into the equation of desirability. Plenty of Americans agree: the campus of Vanderbilt University or Williams College is filled with students whose families are largely indifferent to the price their school charges but keenly interested in the flower beds and the spacious suites and the architecturally distinguished lecture halls those high prices make possible. Of course, given that the rising cost of college has become a significant social problem in the United States in recent years, you can make a strong case that a school ought to be rewarded for being affordable.

The U.S. News rankings turn out to be full of these kinds of implicit ideological choices. One common statistic used to evaluate colleges, for example, is called “graduation rate performance,” which compares a school’s actual graduation rate with its predicted graduation rate given the socioeconomic status and the test scores of its incoming freshman class. It is a measure of the school’s efficacy: it quantifies the impact of a school’s culture and teachers and institutional support mechanisms. Tulane, given the qualifications of the students that it admits, ought to have a graduation rate of eighty-seven per cent; its actual 2009 graduation rate was seventy-three per cent. That shortfall suggests that something is amiss at Tulane. Another common statistic for measuring college quality is “student selectivity.” This reflects variables such as how many of a college’s freshmen were in the top ten per cent of their high-school class, how high their S.A.T. scores were, and what percentage of applicants a college admits. Selectivity quantifies how accomplished students are when they first arrive on campus.

Each of these statistics matters, but for very different reasons. As a society, we probably care more about efficacy: America’s future depends on colleges that make sure the students they admit leave with an education and a degree. If you are a bright high-school senior and you’re thinking about your own future, though, you may well care more about selectivity, because that relates to the prestige of your degree.

There is no right answer to how much weight a ranking system should give to these two competing values. It’s a matter of which educational model you value more—and here, once again, U.S. News makes its position clear. It gives twice as much weight to selectivity as it does to efficacy.

Rankings are not benign. They enshrine very particular ideologies, and, at a time when American higher education is facing a crisis of accessibility and affordability, we have adopted a de-facto standard of college quality that is uninterested in both of those factors. And why? Because a group of magazine analysts in an office building in Washington, D.C., decided twenty years ago to value selectivity over efficacy,