THE RIGHT WAY TO CAPTURE COLLEGE "OPPORTUNITY"

Popular Measures Can Paint the Wrong Picture of Low-Income Student Enrollment

HIGHER EDUCATION may be one of the most important channels through which people can attain improved life outcomes based on their merit rather than family background. If qualified students from lower-income families are underrepresented in higher education, there is potentially a failure not just in equity but in economic efficiency as well.

The question of which colleges and universities lag (or lead) in providing access for low-income students has become a frontline issue in national discussions of educational opportunity. Legislative initiatives such as the bipartisan ASPIRE Act proposed in the U.S. Senate in 2017 would rank institutions based on their percentage of low-income students and impose financial penalties on institutions below a certain ranking. The latest version of the *U.S. News & World Report* "Best Colleges" rankings includes measures of "social mobility." Other news

outlets like the *New York Times* and the *Washington Monthly* have prominently published rankings of colleges based on representation of low-income students while taking editorial positions excoriating (or applauding) individual institutions based on such measures.

Unfortunately, these initiatives ignore a thorny measurement challenge, one that can turn good intentions into penalties for institutions that are actually succeeding in providing opportunities for low-income students and trigger rewards for institutions that are less successful than they might appear. What makes measurement challenging is that different institutions face students whose family income and preparation differ.

Suppose that the University of Maine and the University of Connecticut enroll all students from their respective states who fit their academic standards, as defined by receiving a

by CAROLINE HOXBY and SARAH TURNER

score on a college admissions test that is within the range of most students currently enrolled. Based on the different populations of their states, the University of Maine would draw 22 percent of its students from families with incomes below $40,000, while the University of Connecticut would draw only 10 percent of its students from this income range. The University of Maine would be judged much more favorably by popular measures of "opportunity" and rewarded by proposed accountability systems, while the University of Connecticut would be penalized. However, those rewards and penalties could not result from their differential success in enrolling low-income students since, in the example, all relevant students enroll at each university, regardless of their incomes. The universities would be rewarded based on their circumstances, not their behavior or effort.

More generally, popular measures of "opportunity" confound differences in universities' effort with differences in their circumstances. Specifically, while the measures mean to measure a university's effort to enroll well-qualified low-income students, what they actually measure can largely reflect differences in the pools of students from whom the universities could plausibly draw. The popular measures include a university's share of students who receive federal Pell grants (the "Pell Share"), the share whose family income is in the bottom 20 percent of the national family income distribution (the "Bottom Quintile" measure), and the Intergenerational Mobility ("IGM") measure, which is based on the percentage of enrolled students whose families are in the bottom 20 percent but who themselves end up in the top 20 percent of the national income distribution.

Does a university that does well on these popular measures necessarily have more successful policies for recruiting low-income students than a university that does poorly on these measures? As we show in this analysis, the answer is no.

some universities slated for rewards based on the popular measures actually serve relatively few low-income students from their pool. The reverse is also true: some universities that are slated for penalties based on the popular measures actually serve disproportionately many low-income students from their pool. Thus, measurement matters greatly in this context: judging institutions using flawed measures is likely to produce unintended outcomes because they often give the wrong answer.

Second, we propose a sound measure of a university's success in providing opportunities to low-income students. Specifically, we show how to construct a university's "relevant pool"—the pool of students from which it could plausibly draw based on its academic mission and geographic location. We illustrate how to compare a university's students to its relevant pool, and we demonstrate that such comparisons are highly informative—to show not just how the university serves low-income students but how it serves all students.

The Pell, Bottom Quintile, and IGM measures could be regarded as reasonable proxies for universities' effort in recruiting low-income students if, when tested, they proved to be closely aligned with measures based on universities' relevant pools. If they are not closely aligned, then they must be measuring something different from what is intended. Even worse—because "top performers" and "bottom performers" receive most of the attention— is if the popular measures identify top performance as bottom performance and vice versa.

In addition to our two main aims, we discuss the IGM measure in a bit of detail because it is apparently misunderstood. We show that it shares the flaws of the Bottom Quintile measure but also has additional flaws that lead it to punish universities that face relevant pools with high levels of income *equality*. We conclude with a broad

POPULAR MEASURES OF "OPPORTUNITY" confound differences in universities' effort with differences in their circumstances.

To be clear, we are not criticizing the intentions behind the efforts to measure the success of institutions in providing opportunities for low-income students. Rather, we are attempting to give higher-education leaders the understanding and tools needed to conduct self-evaluation that is likely to further those good intentions.

What This Analysis Does (And Does Not) Do

Our analysis has two main aims. First, we provide a proof by contradiction. That is, we demonstrate that

discussion of how universities can evaluate themselves in a sound manner that could allow them to improve on goals of providing opportunity.

Having said what we attempt, it is worth saying what we do not attempt. Although we suggest measures by which schools could judge whether they are accomplishing their missions, we do not seek to define those missions. These differ in terms of the backgrounds and preparation of the students served. For instance, Berea College states that its mission is: "To provide an educational opportunity for students of all races, primarily from Appalachia, who

have great promise and limited economic resources." This statement defines Berea's relevant pool (all races, primarily Appalachian, of great promise) and its income representation goal (disproportionate emphasis on low-income students). The measure we propose would allow Berea to judge itself against its own mission, but we do not propose to impose a mission on Berea.

Precisely because we do not want to impose missions on universities, we use examples drawn from states' most selective or "flagship" public universities for our proof by contradiction and our illustration of a sound way to measure opportunity. We use them because their key undergraduate mission and constraints are a matter of public record—largely to educate well-prepared students from their own state. Thus, we know approximately how they would define their relevant pools, and we can

between income and preparation, how can we measure an institution's enrollment of students from across the income distribution?

Constructing the "Relevant Pool"

As examples, we use the main campuses of the flagship universities of Connecticut (Storrs), Maine (Orono), Illinois (Urbana-Champaign), Montana (Missoula), New Mexico (Albuquerque), and Wisconsin (Madison). We chose these universities because their relevant pools are distinct in ways that affect measurement. Since the main contributions of this analysis are the proof by contradiction and demonstration that sound measures are possible, we needed to choose interesting universities, not average ones. Our aim is certainly *not* to rank all universities—

WE ILLUSTRATE HOW TO COMPARE A UNIVERSITY'S STUDENTS TO ITS RELEVANT POOL, and demonstrate that such comparisons show not just how the university serves low-income students, but how it serves all students.

construct those pools with a fair degree of confidence. However, the measurement issues we confront apply just as much to non-flagship institutions that are more or less selective and public, nonprofit, or for-profit. Even universities like Harvard and Stanford, which claim to recruit students nationally, in fact have relevant pools that differ substantially owing to strong geographical skews.

Even though flagships' missions and constraints are quite public and the pools we construct for them are grounded in empirical evidence about their behavior, we emphasize that our assumptions are meant only to facilitate illustration. They do not preclude a university specifying alternative parameters.

Furthermore, we do not attempt in this analysis to answer fundamental questions such as why students' preparation varies with family background, why different institutions have curricula and resources designed to serve students with different levels of preparation, and why students often prefer more proximate institutions even when not constrained to attend them. These questions are of absorbing interest to us and other economists of higher education, but we stick here to a simpler question: Given the curricula offered by various institutions (which implicitly constrain the students for whom their offerings generate a high return), given the legal and market conditions under which institutions operate (which affect how attractive they are to out-of-state or otherwise distant students), and given the correlation

indeed, we should refrain from doing so because it is a university's responsibility (and not our right) to define its mission and, thereby, its relevant pool.

We employ measures from de-identified tax data and the population of SAT and ACT college-entrance test-takers from the high school class of 2008. These data, which we also used in the example above, enable us to construct each university's relevant pool by including all students from the state whose scores on either college assessment put them in their flagship's "core" preparation range. Universities report these score ranges—the 25th and 75th percentiles of their students' scores—to the federal Department of Education and to college guides. While selective universities consider multiple indicators of preparation and most practice holistic admissions, these core ranges efficiently summarize institutions' academic standards and are broadly comparable, more so than, for instance, grades.

The data in Figure 1 reveal just how much the relevant pool's income distribution differs for each of these universities. The 20th, 40th, 60th, and 80th percentiles of each income distribution are marked to facilitate comparisons. For instance, compare the University of Connecticut and University of Maine distributions. Maine's 20th percentile income is much lower than Connecticut's 20th percentile income. In fact, Connecticut's 20th percentile is approximately the same as Maine's 40th percentile, and Connecticut's 40th percentile is midway between Maine's 60th and 80th percentiles.

The Illinois-Montana comparison of relevant pools generates similar insights. The 20th percentile for Illinois is higher than the 40th percentile for Montana, and the Illinois 40th percentile is between Montana's 60th and 80th percentiles. Clearly, if one sets any low-income threshold based on a national distribution, as the Pell and Bottom Quintile measures do, a larger share of Maine's or Montana's relevant pool will fall below it. These comparisons illustrate how universities could be penalized for facing higher income distributions (Connecticut, Illinois) or rewarded for facing lower ones (Maine, Montana, New Mexico).

The University of Wisconsin's relevant pool is interesting because the state of Wisconsin has a relatively equal income distribution. (Notice that although Wisconsin's 40th, 60th, and 80th percentiles are well below those of Connecticut and Illinois, Wisconsin's 20th percentile is about t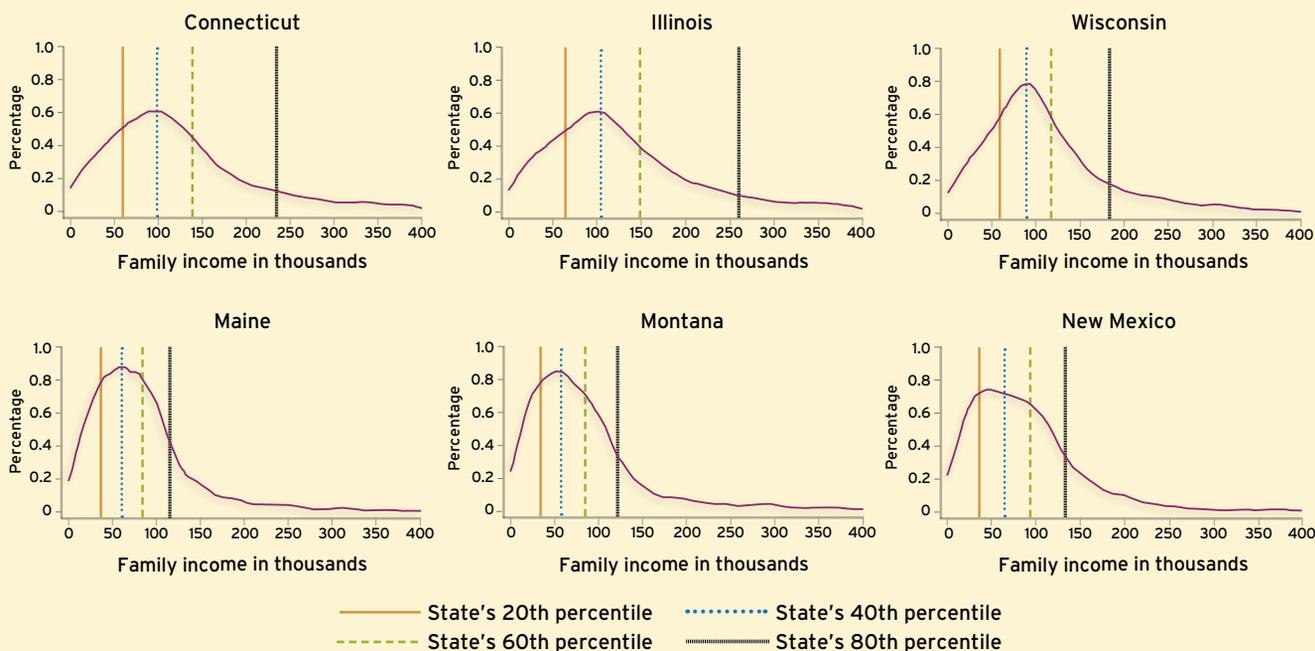he same as theirs.) Wisconsin's income equality translates into relatively few students with very low incomes by national standards. Thus, Wisconsin's relatively equal income distribution—which is probably *good* for disadvantaged students—generates penalties for the university when it is evaluated on Bottom Quintile or Pell measures. Ironically, the university would look better if Wisconsin had more unequal incomes—as does California, for example. Policymakers probably do not mean to penalize universities for their pools' income equality or reward them for inequality.

A Better Measuring Stick

Incorporating information on each university's relevant pool addresses the measurement challenge. In the context of our examples, we examine how universities' in-state enrolled students' income distributions fit into the income distributions of their relevant pools.

Different Universities Have Distinct Relevant Pools (Figure 1)

If flagship state universities served every student in their state who meets their academic preparation standards, the composition of their student bodies would vary widely. For example, the University of Maine would draw 22 percent of students from families with incomes below $40,000 compared to 10 percent at the University of Connecticut.


State's 20th percentile ········· State's 40th percentile
----- State's 60th percentile ▬▬▬ State's 80th percentile

NOTE: The curves on each graph show the percentage of students within each university's relevant pool from families with incomes in every thousand-dollar interval.

SOURCE: Authors' calculations based on 2008 income-tax data

Relevant-Pool Measures Reveal Student Representation at All Incomes (Figure 2)

While the universe of potential enrollees in Connecticut is more affluent than its counterpart in Maine, low-income students in Connecticut are overrepresented at their flagship state university.
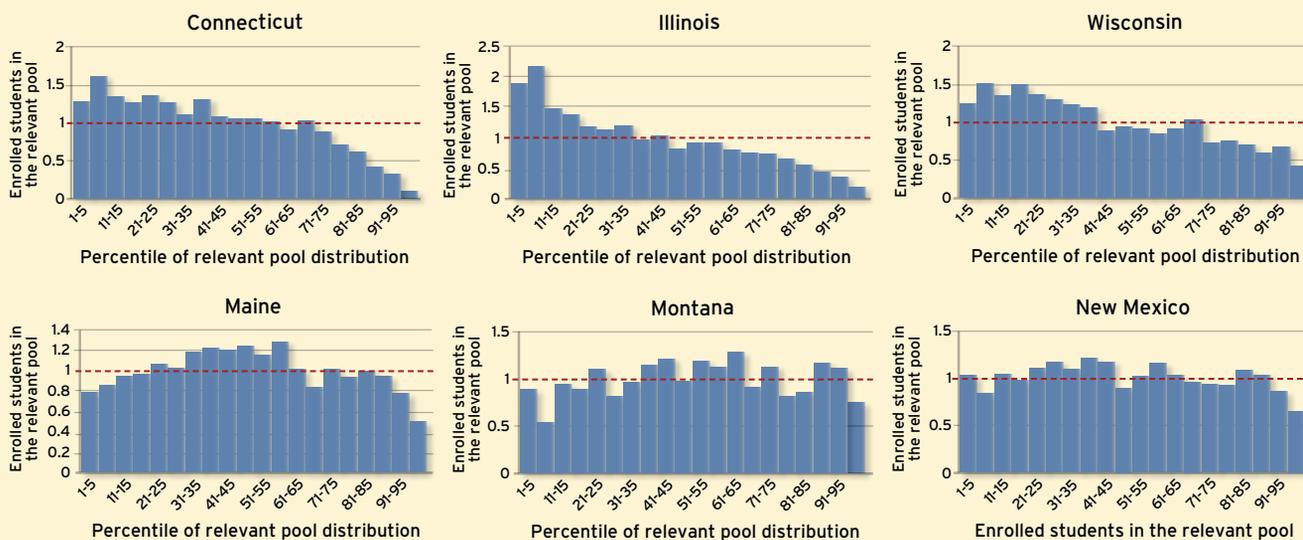
NOTE: Each chart shows the distribution of family income of the students at a state's flagship university fit into the distribution of family income of the students in the university's relevant pool. The bars are based on 5-percentile ranges of the relevant pool's income distribution. The horizontal line set at 1 is an equal representation marker.

SOURCE: Authors' calculations

Specifically, we divide the relevant pool into 5-percentile-wide bins and compute what percentage of each university's in-state students fall into each (see Figure 2). If the university is enrolling prepared students of all incomes equally, each bin will contain 5 percent of students. We divide the bin's percentage by 5 so that the number 1 is a useful marker on the "measuring stick." For instance, if the height of the 21st to 25th percentile bin is 1, then the university's representation of enrolled students from the 21st to 25th percentiles is exactly the same as their representation in the relevant pool. If the height is 1.5, the university's representation of enrolled students is 50 percent greater than their representation in the relevant pool. If the height is 0.5, its representation is 50 percent lower.

Although 1 is a useful marker, it is just a marker—not a mission we impose on schools. For instance, Berea College and many other universities—public and private—that have a mission to serve disadvantaged students especially might want to see numbers above 1 for low-income students. A flagship university might be unconcerned if its numbers were lower than 1 for high-income

students—especially if the school were aware that it offered opportunities to high-income students but that some chose to attend private universities with comparable curricula at their own expense (saving taxpayers' money, thereby, for potential reallocation to needier students).

At each of the universities of Connecticut, Illinois, and Wisconsin, the height of the bars is consistently above 1 for enrolled students from low-income backgrounds—up through at least the 40th percentile of the relevant pool's income distribution. In other words, these universities recruit low-income students sufficiently effectively that such students' representation is disproportionately large. In contrast, the height of the bars is consistently below 1 for enrolled students from low-income backgrounds at the universities of Maine, Montana, and New Mexico, indicating that low-income students' representation is disproportionately small.

At the universities of Connecticut, Illinois, and Wisconsin, the height of the bars for middle-income students is about 1, indicating that their representation is similar to their representation in the relevant pool. At the universities of Maine, Montana, and New Mexico, the
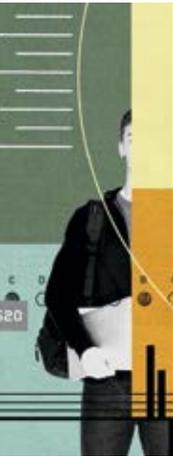
height of the bars for middle-income students is consistently above 1, indicating that middle-income students' representation is disproportionately large.

At all six universities, the height of the bars tends to be

year-to-year changes, particularly for small schools (see "Randomly Accountable," *features,* spring 2002). With that in mind, a university might want to set a more flexible target, such as by using moving averages.

RELEVANT-POOL COMPARISONS ILLUSTRATE HOW UNIVERSITIES COULD BE PENALIZED for facing higher income distributions (Connecticut, Illinois) or rewarded for facing lower ones (Maine, Montana, New Mexico).

below 1 for high-income students. Although we cannot be sure, this is probably not due to the flagships' failing to provide upper-income students with opportunities but, rather, those students choosing to attend private universities at their own expense.

These examples show key advantages of our method:

1. Unlike threshold-based metrics like the Pell, Bottom Quintile, and IGM measures, our method shows how each university is enrolling students across the whole of its relevant pool's income distribution. This comprehensiveness allows observers to take in the entire picture or focus on whatever part of the income distribution interests them.

2. Our method provides a measuring stick but does not impose a mission on a university. A university can choose its own targets across the income distribution—which may include enrolling low-income students disproportionately.

3. Our method does not encourage perverse behavior such as neglecting students just above an arbitrary income threshold. This is unlike the popular measures that make such students—who may need substantial financial aid and encouragement—fail to count toward a university's ranking. Moreover, when—as in proposed federal legislation—all universities face rewards and penalties based on the same threshold, there is increased likelihood of an "arms race" to enroll threshold-eligible students (for example, Pell students), exacerbating any tendency to focus aid on them at the expense of other modest-income students.

Two additional comments are in order. First, because the composition of a university's applicants and relevant pool inevitably vary from year to year, its leaders might compromise academic standards if they try to achieve the same income representation targets every year. Indeed, a well-known result presented by Thomas Kane and Douglas Staiger in their research on K-12 accountability is that it is important to avoid over-interpretation of

Second, a university also might wish to assess the extent to which it has exhausted the pool of relevant students or—alternatively—"left some on the table." It could do this by considering the raw number of students in each 5-percent income bin, rather than percentages, both in the school's relevant pool and the students who enroll. By dividing the number of enrolled students by the number of relevant-pool students in each bin, school leaders could determine the university's "utilization rate," or the size of its class relative to its market, at all 20 levels of the applicable income distribution.

Suppose that the University of Wyoming were assessing whether it had exhausted its pool. Since it is the only baccalaureate-granting public university in a state that has only one (tiny) private baccalaureate-granting institution, it might look for utilization rates fairly close to one as indicating exhaustion. (A rate of 1 would be over-exhaustion because some Wyoming students attend out-of-state.) In contrast, a utilization rate that would indicate exhaustion for the University of California-Berkeley or the University of California-Los Angeles would be well below 0.5. These two flagships share the same relevant pool and California, moreover, contains numerous other public and private institutions whose relevant pools overlap with the flagships' pools. And this is before accounting for California students' tendency to attend out-of-state schools.

Comparing Relevant-Pool and Threshold Rankings

While threshold-based measures and rankings based on them are fundamentally flawed, it is nevertheless informative to compare such rankings with those based on universities' relevant pools—if only to illustrate the magnitude of mis-measurement.

In order to do that, we ranked flagship universities in all 50 states on the shares of their enrolled students whose family incomes fell below the 20th and 40th percentiles of the relevant pool distribution. We then also ranked the universities using the two dominant threshold measures:

the Pell Share and Bottom Quintile approach. Our ranking is such that 1 is the "best" at enrolling low-income students according to the measure being used, and 50 is the "worst." Table 1 shows where the six universities featured in this analysis rank using different measures.

The University of Illinois is ranked 2nd on both of the relevant-pool-based measures. However, it ranked 36th on the Pell measure and 26th on the Bottom Quintile measure. Similarly, the universities of Connecticut and Wisconsin are among the several best on the relevant-pool-based measures. However, they are in the bottom 15 schools on the Pell and Bottom Quintile measures.

Despite the fact that low-income students are well represented in relation to the relevant pools at these three universities, policies based on the popular measures would punish them in various ways, either because the mean income of their relevant pools is high, their income distribution is relatively equal, or both.

The University of Montana is ranked 47th and 40th, respectively, on the first and second relevant-pool-based measures. However, it is ranked 3rd on the Pell measure and 7th on the Bottom Quintile measure. Similarly, the universities of Maine and New Mexico rank among the bottom 15 schools on the relevant-pool-based measures but rank in the top 5 on the Pell and Bottom Quintile measures. Therefore, despite their own states' low-income students being underrepresented at these universities, policies based on the popular measures would reward the universities of Montana, Maine, and New Mexico because they face relevant pools with low incomes, relatively unequal income distributions, or both.

We selected six universities to demonstrate how measurement matters. To be sure, there are universities that would rank similarly regardless of whether we use the relevant pool, Pell, or Bottom Quintile measures. This is because some universities happen to have income-achievement distributions in their relevant pools that are very similar to the national distribution. But that's no reason to endorse the Pell or Bottom Quintile measures. Observing that the measure does not affect some universities' rankings much is akin to observing that it

would not matter how we measured height if everyone were equally tall.

Measuring Mobility along with Opportunity

The increasingly popular Intergenerational Mobility (IGM) measure has received ample attention, including favorable coverage in the *New York Times,* and has been presented as a measure of the effect of universities on the economic success of low-income students. The IGM is calculated by multiplying a university's Bottom Quintile measure by the estimated probability that the university's students from the national bottom income quintile end up, as adults, in the national top quintile.

The IGM suffers from two problems. First, since two

DIFFERENT MEASURES PRODUCE DIFFERENT RANKINGS (Table 1)

Low-income students are well-represented in relation to the relevant pools at flagship universities in Connecticut, Illinois, and Wisconsin, but policies that rank universities on threshold-based measures such as the Pell and Bottom Quintile would punish them.

| | RANK ON | | | |
FLAGSHIP UNIVERSITY	SHARE BELOW 20TH PERCENTILE OF RELEVANT POOL	SHARE BELOW 40TH PERCENTILE OF RELEVANT POOL	PELL SHARE	BOTTOM QUINTILE MEASURE
ILLINOIS	2	2	36	26
WISCONSIN	6	4	48	36
CONNECTICUT	7	5	35	41
NEW MEXICO	34	30	2	1
MAINE	42	34	5	3
MONTANA	47	40	3	7

SOURCE: Authors' calculations

incomes, relatively unequal income distributions, or both.

thirds of the variation in the IGM comes from variation in the Bottom Quintile measure, these measures share the same flaws. Second, the mobility measure rewards relevant pools with extreme income inequality and penalizes those where incomes are more equal.

Looking again at the University of Wisconsin, the state's unusually equal income distribution means that it has comparatively few adults in both the bottom and the top national income quintiles. Thus, University of Wisconsin students who stay in the state are disproportionately unlikely to end up in the national top quintile, regardless of

the income with which they grew up. The mobility measure penalizes the state's income equality twice: once through the Bottom Quintile measure (since the state is too equal to have many people in the bottom quintile) and again through the mobility measure (since the state is too equal to have many people in the top quintile). It is no surprise that the University of Wisconsin ranks 22nd out of the 25 highly selective public colleges for which this measure is reported via the *New York Times*. Ironically, the University of Wisconsin would look better on the ranking if the state of Wisconsin had less equal incomes.

For universities in states that face unusually unequal distributions, the situation is reversed. For instance, the California flagships face a relevant pool with large percent-

find that the easiest way to attain a better ranking is to deviate substantially from its educational mission.

While the examples of this study illustrate the naiveté of applying national benchmarks to universities in vastly different circumstances, we have purposely avoided ranking all institutions because that would require us to assert the relevant pool—and thus the mission—of thousands of institutions of higher education in the United States. This is not our right. In contrast, it is not merely the right but the duty of the leaders and stakeholders of each university to define its mission and, thereby, its relevant pool. When they acquiesce to demands to use the popular measures simply because they are trendy, leaders and stakeholders forgo crucial conversations about mission

UNIVERSITIES LIKE FLAGSHIP SCHOOLS IN ILLINOIS, CONNECTICUT, AND WISCONSIN RANK AMONG THE SEVERAL BEST on the relevant-pool-based measures. However, they are in the bottom 15 schools on the Pell and Bottom Quintile measures.

ages of students at both the very top and very bottom of the income distribution. Thus, the IGM and other popular measures "reward" them for their state's income inequality.

Measurement Matters

It is precisely because selective universities can engage students in intense learning that they have the capacity to transform lives. Indeed, there would be little point in urging universities to offer opportunities to students of all backgrounds if those opportunities were not valuable. Thus, it would seem unwise to coerce universities to neglect their educational missions.

The mis-measurement of "opportunity" for low-income students, which is all too common in Pell Share, Bottom Quintile, and IGM measures, produces perverse incentives. As we have shown, differences in the income-achievement distributions of universities' relevant pools can produce penalties and rewards that are not only unintended but even the inverse of what was intended. In addition, when encountering strong policy incentives or public pressure to improve on these flawed measures, institutions could respond in ways that undermine their missions, such as by enrolling less-prepared students who meet the Pell or Bottom Quintile threshold even when there are much better prepared students just above the threshold, or by substituting out-of-state students who meet the threshold for in-state students. A university that pursues the flawed measures, which confound its behavior with its circumstances, may

and constraints. Deliberating actively and often on the university's relevant pool is a prerequisite to assessing the extent to which a university is succeeding in attracting students from across the income distribution.

A university that used our proposed relevant-pool-based measure would not find a conflict between pursuing its mission and providing opportunities to students regardless of background. In prior research, we have shown that the institutions that moved earliest toward enrolling students based on merit rather than background find themselves to be the most influential and financially capable today. Moreover, because there are more high-achieving low-income students who could benefit by attending selective institutions than currently enroll, a university could probably gain from rigorous, data-driven self-examination—deciding what its relevant pool is, comparing its students to the pool, and drawing lessons for recruitment, admissions, and financial aid. If universities use methods that measure what is *intended,* they can further both equity and excellence simultaneously.

Caroline Hoxby is the Scott and Donya Bommer Professor in Economics at Stanford University and a senior fellow of the Hoover Institution. Sarah Turner is University Professor of Economics and Education and Souder Family Professor at the University of Virginia. This article is drawn from "Measuring Opportunity in U.S. Education," which is available as a working paper from the National Bureau of Economic Research.