



# Developing a Multimetric Accountability System for Postsecondary Institutions

*Sandy Baum, Erica Blom, and Jason Cohn*

*June 2022*

Without strong safeguards to hold postsecondary institutions accountable for student outcomes, students and taxpayers will continue to pay a high cost for investments in college education that do not pay off. There will always be students who do not meet their postsecondary goals, but many bad outcomes could be prevented if the federal government held programs and institutions to higher standards.

We examine the impact of using a system based on multiple metrics that requires institutions to pass three out of four thresholds for student loan default, student loan repayment, program completion, and postcollege earnings. Currently, a very high loan default rate is the only student outcome that disqualifies institutions from the federal student aid system. Relying on multiple metrics diminishes the risk of institutions manipulating their outcomes and requires satisfactory performance in more than one area, while allowing flexibility for differing programs, missions, and circumstances.

We do not suggest that the metrics we discuss are the “right” ones. Rather, we use reasonable metrics to illustrate the importance of a range of choices, including weighting default rates by the share of students borrowing; focusing on the share of debt retired as opposed to the share of students retiring any debt; setting different completion thresholds for four-year, two-year, and less-than-two-year institutions; and monitoring indicators of the distribution of earnings rather than just graduates’ average earnings, with different thresholds for different types of institutions. See [our related report](#) on this subject for our analysis of other constructive adjustments (Baum, Blom, and Cohn 2022). These adjustments can have a major impact on which institutions pass and which institutions fail the accountability test. We conclude that minimum standards should apply to all institutions regardless of their student bodies, but thresholds, particularly for earnings and completion rates, should differ by program length.

Using multiple metrics diminishes the risk of inappropriately punishing institutions because of the fit between their programs and students and the standards imposed. If institutions must meet only a subset of the thresholds, they can avoid some of the vulnerability associated with particular students and program characteristics. No metric or set of metrics will perfectly measure institutional quality. But adhering to basic policy design principles can lead to an effective system that reduces the number of students bringing their federal financial aid to programs and institutions that are unlikely to help them achieve their goals.

## Design Principles

Adhering to several principles in addition to including multiple metrics and providing flexibility will create a stronger, more effective accountability system. Standards should be simple enough for institutions and possibly students to understand them, and adequate data must be available to implement the standards. Analyzing data on the impact of different standards makes the significance of several other principles clear. (See the associated report for further explanation of these principles.)

**Standards should apply to institutions in all sectors but recognize the particular risks of the for-profit structure, as well as limited resources at some institutions, particularly community colleges and minority-serving institutions.** In all sectors, there is a wide range of student outcomes, and all institutions should be held accountable. But differences in governance structures and the historical record point to disproportionate problems in the for-profit sector. And penalizing institutions that are severely handicapped by funding shortfalls is unlikely to help students.

**Standards should distinguish between four-year, two-year, and less-than-two-year programs and institutions.** Devising and implementing lower standards for institutions that enroll large numbers of students at high risk of noncompletion is complex and risks a failure to ensure high standards for all students. We suggest varying thresholds only by program length.

Despite attempts to estimate the investment students make (usually net of forgone earnings) and the prices they have paid, estimates of rates of return are imprecise. As a pragmatic compromise, the threshold for acceptable earnings should be lower for short-term programs than for bachelor's degrees and should be higher for advanced degree programs. The time and money required for longer programs requires higher earnings to generate a strong rate of return. A reasonable approach is to compare earnings for graduates of programs offering similar credentials.

**Standards should consider the distribution of student outcomes, not just the average outcome.** Standards could, for example, focus on the share of students earning more than the average high school graduate (or some other minimum threshold), rather than focus on average earnings relative to the threshold. High average earnings could reflect a few graduates with very high earnings rather than typical graduates doing well in the labor market. The goal should be to ensure that few students end up unable to earn a living wage—or not having benefited financially from their investment in postsecondary education.

**Standards should account for geographic and cyclical variation.** It is possible to adjust program graduates' earnings for earnings levels in their areas of residence at the time they enrolled in college. Although postcollege earnings depend on where students live after they leave school, basing adjustments on precollege location provides a better measure of the earnings boost resulting from postsecondary education, because it allows comparison with a reasonable estimate of what earnings would have been had students not attended college.

To avoid a disproportionate number of failures when the labor market is weak and passing marks when the labor market is strong, earnings metrics should be based on multiple years of earnings or multiple cohorts. Adjusting the threshold based on labor market conditions can also mitigate this problem.

**Where appropriate, standards should include all enrollees, not only graduates.** It may not be appropriate to include noncompleters when evaluating earnings because they do not experience the full benefits of the education. But that should not exclude them from the loan metrics, because many students borrowing for a program they do not complete indicates a real problem.

## What Difference Does the Choice of Standards Make?

Based on these design principles, we use College Scorecard data to explore potential thresholds for four student outcomes: default, loan repayment, completion, and earnings. To anchor the discussion, we use the example of thresholds set at levels where 5 percent of students would attend institutions failing the standard. This threshold is low enough not to capture too many institutions but is high enough to make a meaningful contribution to improving the poorest performances. We do not argue for relative metrics, with the threshold changing as overall performance changes. Rather, current performance can guide the choice of initial thresholds, which should be fixed at one level.

### Loan Default: Share of Borrowers or Share of Students?

The current regulatory system relies on cohort default rates (CDRs), measuring the share of borrowers entering repayment who default within a specified period. But the system could instead focus on the share of all students—including those who do not borrow—who default, focusing on the share of students who struggle with debt, rather than just the share of borrowers who struggle.

Focusing on all students gives an indication of the share of students for whom borrowing for education has not worked out well. Adjusting the default rate for the share of students at an institution who borrow—attaching a less favorable score to an institution where most students have debt than to an institution with the same share of borrowers defaulting but a lower share with debt—gives a better indication of this outcome. The problem is greater if a large share of all students struggles with debt than if a large share of the small fraction of students who borrowed struggles with debt.

Because a relatively small share of public two-year college students borrows, using the CDR adjusted for the share of students borrowing moves public two-year colleges from the highest default rate to among the lowest default rates (table 1).

**TABLE 1**  
**Average Cohort Default Rates**

	Average CDR among borrowers	Average CDR among all students
For-profit two-year	13%	9%
For-profit four-year	12%	7%
For-profit less-than-two-year	14%	9%
Nonprofit two-year	13%	11%
Nonprofit four-year	6%	4%
Nonprofit less-than-two-year	13%	7%
Public two-year	16%	4%
Public four-year	6%	3%
Public less-than-two-year	13%	8%
<b>Total</b>	<b>8%</b>	<b>4%</b>

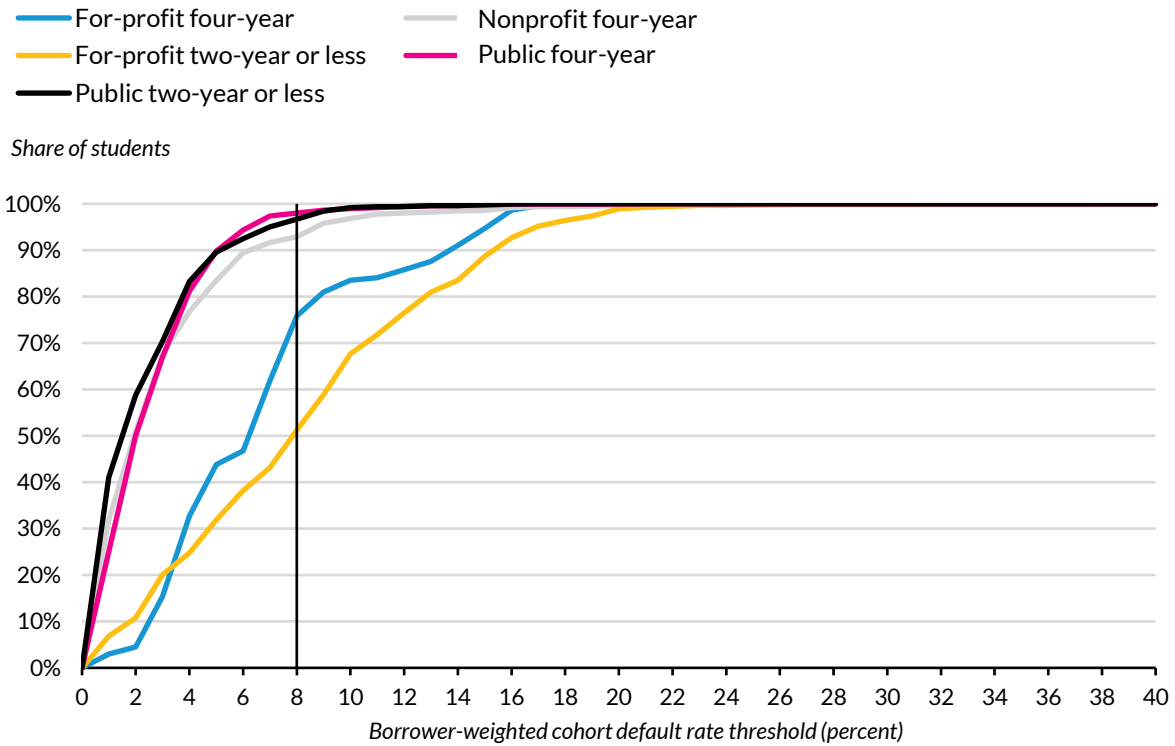
**Source:** Authors' calculations using 2019–20 College Scorecard data.

**Notes:** CDR = cohort default rate. Default rates include Stafford loans for both undergraduate and graduate students but exclude PLUS loans for parents and graduate students. The CDR among all students is calculated for each institution by multiplying the CDR (share of borrowers defaulting) by the share of students borrowing. Averages for sectors are weighted by the share of borrowers enrolled in each institution.

Figure 1 shows the share of students in each sector attending institutions that would pass this CDR metric at each potential threshold. If the threshold were a default rate for all students of just over 8 percent, institutions serving roughly 5 percent of students would fail. At this threshold, very few students at public and private nonprofit four-year institutions would be affected. Public two-year colleges do well on this metric—roughly 10 percent of students at failing institutions attend public two-year colleges (table 2)—reflecting that few of their students borrow. (Focusing only on the share of borrowers defaulting, public two-year colleges fare far worse, with 63 percent of students at failing institutions enrolled in this sector.) For-profits make up about half of students at failing institutions, with nearly half of students at for-profit two-year institutions attending failing institutions.

FIGURE 1

Share of Students Attending Institutions Passing the Adjusted Cohort Default Rate Metric  
Thresholds from 0 to 40 percent



URBAN INSTITUTE

Source: Authors' calculations using 2019-20 College Scorecard data and the Integrated Postsecondary Education Data System.

Note: The line at 8 percent corresponds to the adjusted default rate at which 5 percent of students would attend failing institutions.

TABLE 2

**Adjusted Cohort Default Rate Metric***Five percent of students attend failing institutions*

	Students attending failing institutions	Total students	Share	Failing institutions	Total institutions	Share	Failure threshold
For-profit two-year	124,196	266,163	47%	145	404	36%	8%
For-profit four-year	191,075	793,165	24%	57	170	34%	8%
For-profit less-than-two-year	91,087	204,355	45%	277	927	30%	8%
Nonprofit two-year	23,429	44,852	52%	17	118	14%	8%
Nonprofit four-year	156,896	3,613,508	4%	132	1,297	10%	8%
Nonprofit less-than-two-year	1,137	7,030	16%	8	35	23%	8%
Public two-year	83,818	3,580,631	2%	31	933	3%	8%
Public four-year	133,507	7,615,539	2%	38	683	6%	8%
Public less-than-two-year	5,271	42,749	12%	50	218	23%	8%
<b>Total</b>	<b>810,416</b>	<b>16,167,992</b>	<b>5%</b>	<b>755</b>	<b>4,785</b>	<b>16%</b>	<b>8%</b>

Source: Authors' calculations using 2019–20 College Scorecard data and the Integrated Postsecondary Education Data System.

### Repayment Rates: Share of Borrowers or Share of Dollars?

Institutions may attempt to manipulate the default rate by pushing borrowers who are not making their loan payments into forbearance until the default measurement window closes. Moreover, as more borrowers enroll in income-driven repayment plans, more of those who cannot pay can remain in good standing, avoiding delinquency and default. Adding a metric related to successful loan repayment to complement the student-based default rate metric captures debts disproportionate to incomes, even when borrowers are delinquent but not in default or are in good standing because of income-driven repayment, despite not retiring their debts. The repayment rate metric is difficult for an institution to manipulate.

In contrast to the standard repayment rate defined by the US Department of Education in the College Scorecard data, which measures the share of borrowers in a cohort who reduce their loan principal by at least one dollar within a specified number of years, dollar-based repayment rates reflect the share of loan dollars outstanding, relative to the amount originally borrowed. This metric is sensitive to differences in repayment success, rather than giving each borrower a yes-or-no rating. It differentiates between institutions where most students manage repayment successfully and those where this success is minimal.

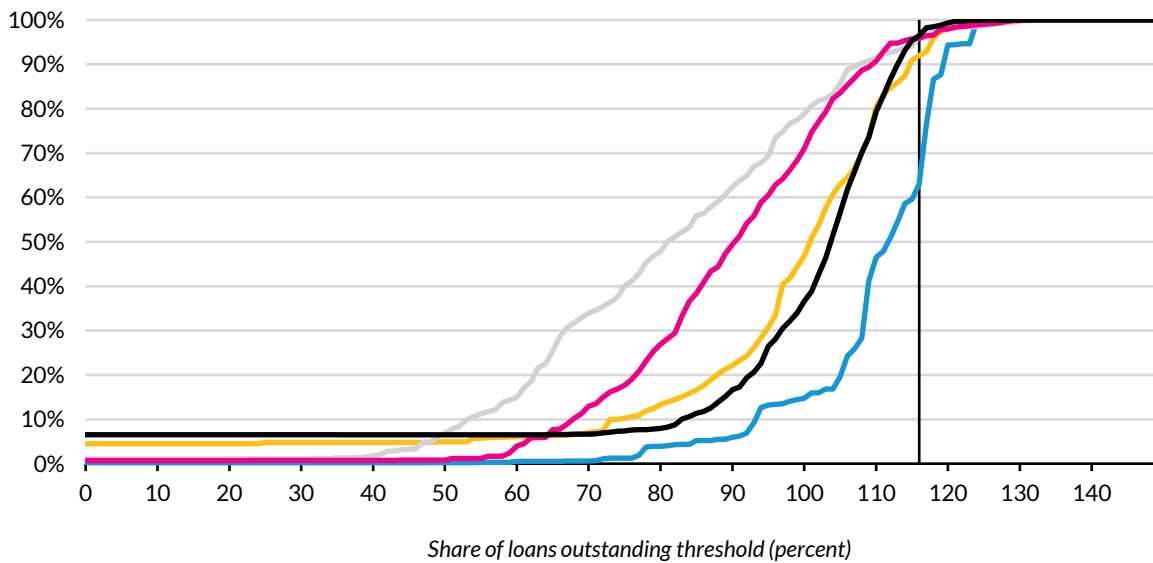
The 95th percentile of the share of loan dollars outstanding is 117 percent (5 percent of students attend institutions where total outstanding debt is at least 17 percent higher than the original amount borrowed five years into repayment). One-third of undergraduate students in the four-year for-profit sector attend institutions where the outstanding loan balance after five years is more than 117 percent of the amount borrowed. The shares of students attending such institutions in other sectors are low, but 42 percent of all undergraduate students attend institutions where the dollars outstanding are greater than the initial balance after five years. Figure 2 shows the share of students in each sector attending institutions that pass different dollar-based repayment rate metrics.

**FIGURE 2**  
**Share of Undergraduate Students Attending Institutions Passing the Dollar-Based Repayment Metric after Five Years**

*Share of dollars outstanding*

- For-profit four-year
- For-profit two-year or less
- Nonprofit four-year
- Public four-year
- Public two-year or less

*Share of students*



URBAN INSTITUTE

Source: Authors' calculations using 2018–19 College Scorecard data and the Integrated Postsecondary Education Data System.

## Completion Rates

Including completion in accountability is important because it is a meaningful outcome of higher education that applies to all students, not only those who rely on debt financing. Moreover, unlike default, repayment, and earnings measures, completion rates are available quickly. The optimal time

period for measuring completion is debatable, but it is shorter than the time required for students to settle into the labor market or establish their repayment patterns.

**TABLE 3**

**IPEDS Completion Rates of First-Time Full-Time Undergraduate Students at Starting Institution**

	Enrollment-weighted avg.	Institutions			Students		
		P25	Med.	P75	P25	Med.	P75
For-profit two-year	61%	56%	67%	77%	53%	62%	68%
For-profit four-year	35%	32%	49%	67%	23%	34%	42%
For-profit less than two-year	69%	61%	73%	83%	61%	70%	78%
Private nonprofit two-year	60%	36%	65%	83%	55%	64%	71%
Private nonprofit four-year	65%	42%	57%	70%	51%	65%	82%
Private nonprofit less than two-year	71%	58%	70%	80%	58%	77%	80%
Public two-year	27%	22%	29%	39%	20%	26%	32%
Public four-year	57%	34%	46%	59%	44%	57%	73%
Public less than two-year	70%	65%	78%	88%	58%	72%	81%
<b>All institutions</b>	<b>51%</b>						

**Source:** Authors' calculations using 2019–20 College Scorecard data and the Integrated Postsecondary Education Data System.

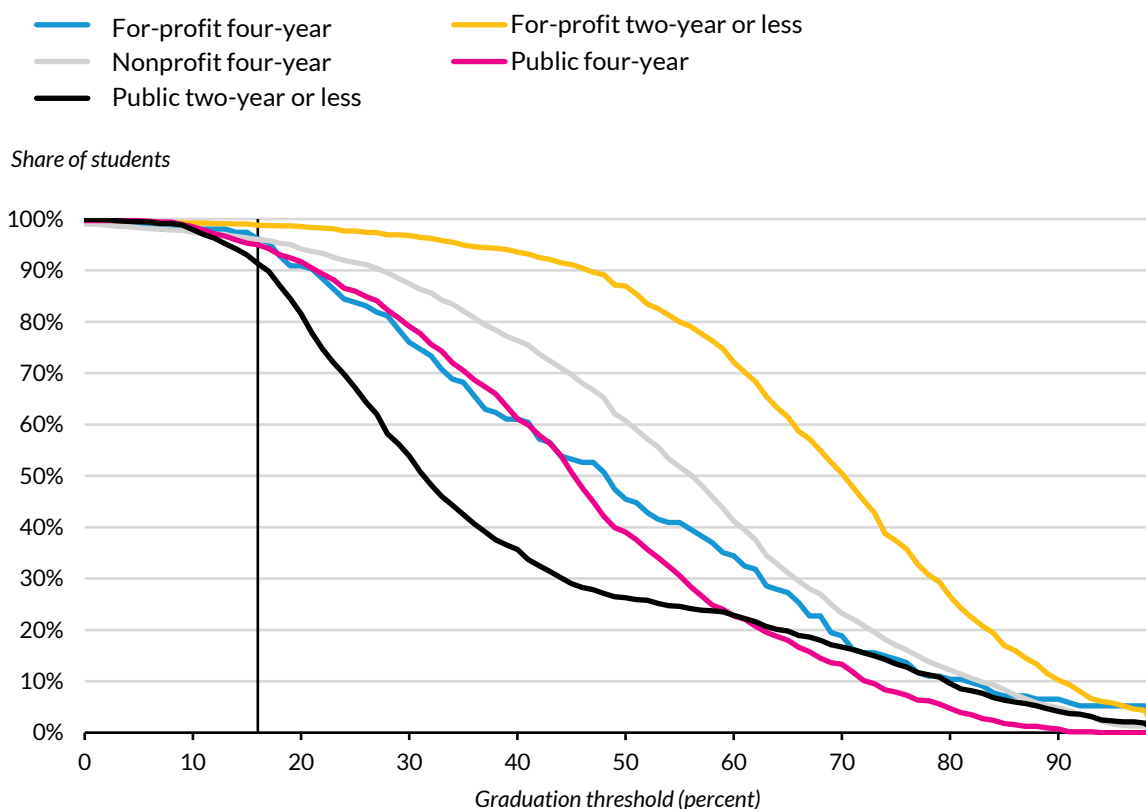
**Notes:** IPEDS = Integrated Postsecondary Education Data System; Med. = median; P25 = 25th percentile; P75 = 75th percentile. Completion is defined as completion at original institution within 150 percent of normal time: six years for four-year institutions, three years for associate degree programs, and varying times for certificate programs.

The shares of students affected by a graduation rate threshold of 16 percent—the level at which 5 percent of students overall would attend failing institutions—range from 7 percent at public two-year-or-less institutions to 1 percent at for-profit two-year-or-less institutions (figure 3). It is, however, problematic to set one completion rate threshold for all types of institutions, given the different barriers to completing four-year degrees in six years, two-year degrees in three years, and short-term certificates in 150 percent of “normal” time. Moreover, completion rates for first-time, full-time students cover different shares of students at different types of institutions. Setting thresholds separately, so that 95 percent of students in each of the three categories of institutions attend schools that meet the criteria, would yield thresholds of 12 percent for two-year institutions, 21 percent for four-year institutions, and 42 percent for less-than-two-year institutions.



FIGURE 3

Share of Students Graduating within 150 Percent of Normal Time



URBAN INSTITUTE

Source: Authors' calculations using 2019–20 College Scorecard data and the Integrated Postsecondary Education Data System.

Note: Completion is defined as completion at the original institution within 150 percent of normal time: six years for four-year institutions, three years for associate degree programs, and varying times for certificate programs.

### Labor Market Outcomes

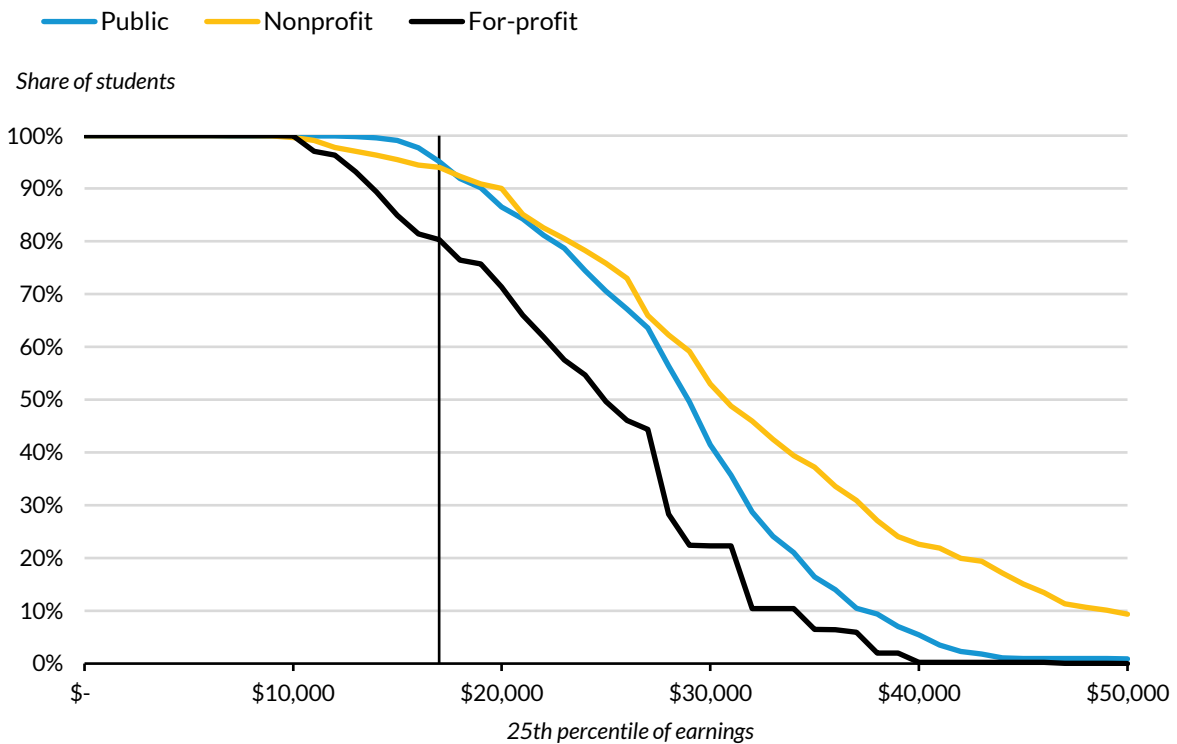
Increased earnings are far from the only valuable outcome of postsecondary education, but virtually all students hope and expect they will be better off financially after completing college than they would have been had they not enrolled. The best measure would involve estimating rates of return. But the difficulty of identifying net price for individual students and of incorporating any measure of forgone wages makes this challenging.

Among the questions critical for developing a reliable earnings metric are how to compare the earnings of students with different types of credentials. A bachelor's degree program where half the graduates earn more than \$40,000 a year is not comparable with a short-term certificate program with this outcome. One reasonable approach is to compare the shares of graduates of similar institutions earning above a certain threshold with thresholds differing in a logical manner across types of institutions.

Rather than comparing earnings with the federal poverty level or average earnings for high school graduates, we set the threshold separately for four-year, two-year, and less-than-two-year institutions at the 25th percentile of earnings for each type of institution. Comparing earnings with the federal poverty level or with average earnings for high school graduates disadvantages short-term programs. Using a cutoff defined by earnings for similar programs or institutions avoids this problem.

The 25th percentile allows for a share of students to pursue nonfinancial measures of success while still requiring most students to earn above a specified threshold. For four-year institutions, the 5th percentile of 25th percentile earnings (using 2014–15 data) is \$16,700 (figure 4). In other words, 5 percent of students attend institutions where more than 25 percent of students earned less than \$16,700 10 years after enrolling. For-profit institutions are less likely than public and private nonprofit four-year institutions to pass this metric. Parallel thresholds would be \$13,000 for two-year institutions and \$8,000 for less-than-two-year institutions.

**FIGURE 4**  
**Share of Students Attending Institutions Where at Least 75 Percent Earn Specified Levels 10 Years after Enrolling**  
*Four-year institutions*



URBAN INSTITUTE

Source: Authors' calculations using 2014–15 College Scorecard data and the Integrated Postsecondary Education Data System.

## A Multimetric System

Below, we use a set of indicators to examine the potential impact of a multimetric system. Requiring that institutions meet more than one benchmark will increase the number that fail. If the metrics are highly correlated, the impact will be small. Allowing institutions to meet, for example, three out of four benchmarks will allow some institutions to pass, even with poor performance in one area. Under the system we examine, 5 percent of four-year students would attend failing institutions—a contrast to the current system, where failure is rare. Nearly 40 percent the failures would be in the for-profit sector, where more than one-third of students would attend failing institutions. (Results for two-year and less-than-two-year institutions are in the accompanying report.) Modifying the threshold would change the share of students and institutions affected.

TABLE 7

**Causes of Failure in a Multimetric System***Share of four-year institutions failing, by sector*

	Public Four-Year		Private Nonprofit Four-Year		Private For-Profit Four-Year		All Four-Year	
	Institutions	Students	Institutions	Students	Institutions	Students	Institutions	Students
<b>Failing two metrics</b>								
Cohort default rate, loan repayment	3.0%	1.3%	1.4%	0.6%	2.8%	0.6%	2.1%	1.0%
Cohort default rate, completion	0.2%	0.0%	0.3%	0.0%	3.7%	2.2%	0.5%	0.2%
Cohort default rate, earnings	0.2%	0.0%	1.2%	0.4%	13.9%	5.9%	1.6%	0.4%
Loan repayment, completion	0.2%	0.1%	0.0%	0.0%	1.9%	14.6%	0.2%	0.8%
Loan repayment, earnings	0.2%	0.1%	0.6%	0.2%	6.5%	1.5%	0.8%	0.2%
Completion, earnings	3.8%	1.2%	0.7%	0.1%	0.9%	0.0%	1.8%	0.8%
<b>Failing three metrics</b>								
Cohort default rate, loan repayment, completion	0.5%	0.1%	0.3%	0.0%	2.8%	8.1%	0.5%	0.5%
Cohort default rate, loan repayment, earnings	0.9%	0.2%	1.8%	0.5%	6.5%	3.3%	1.8%	0.5%
Cohort default rate, completion, earnings	0.0%	0.0%	0.2%	0.7%	1.9%	0.3%	0.2%	0.2%
Loan repayment, completion, earnings	0.5%	0.2%	0.2%	0.1%	0.0%	0.0%	0.3%	0.1%
<b>Failing four metrics</b>								
Cohort default rate, loan repayment, completion, earnings	0.5%	0.1%	0.6%	0.1%	1.9%	0.2%	0.6%	0.1%
<b>Total number failing</b>	<b>62</b>	<b>250,228</b>	<b>73</b>	<b>94,346</b>	<b>46</b>	<b>223,176</b>	<b>181</b>	<b>567,750</b>

Source: Authors' calculations using College Scorecard data and the Integrated Postsecondary Education Data System.

## Conclusion

All students deserve to attend programs and institutions that meet high quality standards. But different program goals, investments of time, and eligibility criteria make it difficult to set standards that apply equally well to all types of programs and institutions. Setting one earnings threshold or one completion rate threshold for all programs and institutions will create a biased system.

The outcomes documented here and in the accompanying report and illustrated in our data visualization indicate that even if we implement effective accountability standards that steer students away from the institutions with the weakest outcomes, much work will remain to ensure students achieve their educational goals. Isolating the institutions with the poorest outcomes requires setting thresholds for default, loan repayment, completion, and postcollege earnings that are below the levels most observers would hope to see for most students. Raising cutoffs to levels indicative of meeting high standards would penalize a large share of postsecondary institutions. Setting standards will provide incentives for improved performance and eliminate programs and institutions that fail the standards. Highlighting outcomes on the proposed metrics should also inspire policymakers and educators to redouble efforts to provide institutions and students the resources and strategies required to support student success.

## Reference

Baum, Sandy, Erica Blom, and Jason Cohn. 2022. *Using Multiple Metrics to Strengthen Institutional Accountability*. Washington, DC: Urban Institute.

## About the Authors

**Sandy Baum** is a nonresident senior fellow in the Center on Education Data and Policy at the Urban Institute and professor emerita of economics at Skidmore College. An expert on higher education finance, she speaks and writes extensively about issues relating to college access, college pricing, student aid policy, student debt, and affordability. Baum earned her BA in sociology from Bryn Mawr College, where she serves on the board of trustees, and earned her PhD in economics from Columbia University.

**Erica Blom** is a research associate in the Center on Education Data and Policy, where she studies higher education policy. Blom received a bachelor's degree in mathematics and political science from Queen's University and a master's degree in economics from Western University. She also earned a doctoral degree in economics from Yale University, where her research focused on students' choices in college major.

**Jason Cohn** is a research analyst in the Center on Education Data and Policy, where he focuses on higher education topics. He graduated from the University of North Carolina at Chapel Hill with

bachelor's degrees in economics and public policy and completed his master's degree in public policy at the George Washington University.

## Acknowledgments

This brief was supported by Arnold Ventures. We are grateful to them and to all our funders, who make it possible for Urban to advance its mission.

The views expressed are those of the authors and should not be attributed to the Urban Institute, its trustees, or its funders. Funders do not determine research findings or the insights and recommendations of Urban experts. Further information on the Urban Institute's funding principles is available at [urban.org/fundingprinciples](https://urban.org/fundingprinciples).



500 L'Enfant Plaza SW  
Washington, DC 20024

[www.urban.org](https://www.urban.org)

### ABOUT THE URBAN INSTITUTE

The nonprofit Urban Institute is a leading research organization dedicated to developing evidence-based insights that improve people's lives and strengthen communities. For 50 years, Urban has been the trusted source for rigorous analysis of complex social and economic issues; strategic advice to policymakers, philanthropists, and practitioners; and new, promising ideas that expand opportunities for all. Our work inspires effective decisions that advance fairness and enhance the well-being of people and places.

Copyright © May 2022. Urban Institute. Permission is granted for reproduction of this file, with attribution to the Urban Institute.