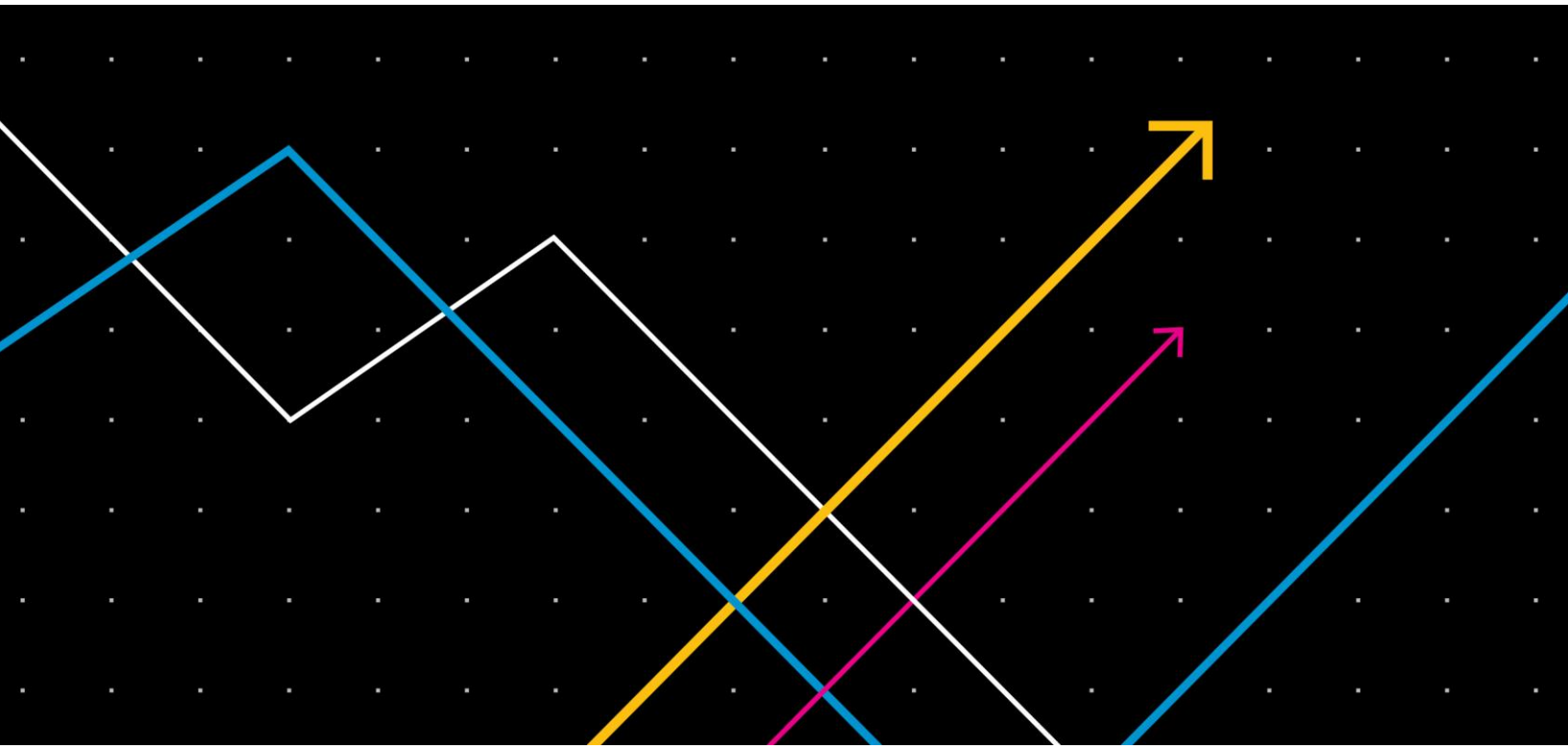


RESEARCH REPORT

# Using Multiple Metrics to Strengthen Institutional Accountability

*Sandy Baum* *Erica Blom* *Jason Cohn*

*March 2022*

URBAN INSTITUTE · ELEVATE · THE · DEBATE


## 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.

# Contents

# Acknowledgments

This report 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.

# Executive Summary

This report sets out principles for developing an accountability system for postsecondary institutions to improve outcomes for students, protecting both them and the investments taxpayers make in their education. We propose a system based on multiple metrics that requires institutions to pass three out of four thresholds based on student loan default, student loan repayment, program completion, and postcollege earnings. 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.

Using illustrative examples of potential metrics, we show the impact 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 and earnings thresholds for four-year, two-year, and less-than-two-year institutions; and adjusting earnings thresholds for geographic differences. 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.

No metric or set of metrics will perfectly measure institutional quality. But adhering to the basic policy design principles set forth in this report will 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.

# Using Multiple Metrics to Strengthen Institutional Accountability

There is broad consensus that the system for holding accountable postsecondary institutions whose students receive federal student aid needs repair. The limited rules in place do not provide adequate protections for students or for taxpayers. Too many institutions with low completion rates, poor employment outcomes, and weak student loan repayment experiences continue to enroll students who depend on federal aid. Many of these students would likely attend colleges offering a better chance of success if the accountability system were stronger.

Researchers, think tanks, and advocates have offered detailed plans for reforming the system. Generally, they choose one or more metrics they argue would be most fair and effective, frequently providing data to indicate the potential impact of the proposed scheme. Limited data availability and uncertainty about the optimal solutions for students facing significant barriers to academic success make it difficult to precisely weigh the merits of the specific metrics proposed.

In this report, we clarify the goals of policy reform and consider underlying principles for effective policy design in this area. In addition to reviewing the metrics of existing studies, we delve into the conceptual grounding for developing metrics. Instead of arguing for a particular metric or set of metrics, we outline the key characteristics of an effective system and the compromises involved in any potential choices.

We use available data to test the likely impact of a set of potential benchmarks, focusing on a system that incorporates multiple metrics assessing different areas of performance, requiring success from more than one perspective but allowing flexibility for institutions whose circumstances create challenges. Some metrics we examine have been suggested by others, and some are new. We look at potential metrics individually, and then we look at their interactions, setting the stage for the implementation of politically feasible, effective measures for protecting future students and taxpayers. Our goal is not to argue for a specific choice of metrics but to ensure policymakers carefully consider the advantages and disadvantages of any metrics that might emerge.

# The Challenges of Universal Standards

A system of thresholds postsecondary institutions should meet is not a ranking system. The idea is not to assign scores to institutions or to be able to list them from best to worst. Rather, the idea is to develop criteria that put institutions on one side of a line or the other or into categories of pass, conditional pass requiring improvement, and fail.

Nonetheless, any set of standards inevitably involves comparisons across institutions. Those that meet the standards are more acceptable than those that do not. So, a first step is to ask whether it is reasonable to have one set of standards that applies to all postsecondary institutions or whether there should be separate standards for each sector, for two-year and four-year institutions or programs, or for institutions enrolling students with different characteristics.

## Setting Standards for Institutions with Different Types of Students

Some participants in the institutional accountability conversation emphasize the importance of comparing programs and institutions only with others offering similar programs to similar students (Cielinski and Pham 2017; Clotfelter 2012; Clotfelter et al. 2013). Their stated goal is to preserve opportunities for students facing the greatest barriers to success. The fear is that if standards are not adjusted for student characteristics, institutions will exclude students facing significant barriers to success. Programs that enroll large numbers of these students will have great difficulty meeting the thresholds. And programs preparing students for occupations that typically have low wages will be unable to operate.

Providing the best possible options for students with weak academic preparation and work and family obligations that limit their flexibility and the time they can devote to school is vital. But equally important is ensuring high standards for all students, not just those with a strong chance of success no matter where and what they study.

## Setting Standards for Institutions with Different Types and Lengths of Programs

In addition to this intractable issue, there are practical considerations involved in weighing the advantages and disadvantages of universal standards. For example, completion rates are a critical metric. Institutions at which very few students complete their programs are problematic regardless of student characteristics. But can we compare completion rates for certificate programs with completion rates for associate degree and bachelor's degree programs?

The standard Integrated Postsecondary Education Data System (IPEDS) completion metric is completion for full-time students at their original institution within 150 percent of the "normal" time (i.e., three years for associate degree programs and six years for bachelor's degree programs), with no clear standard for certificate programs, which vary in length. Although this seems to be a reasonable way to adjust metrics by program level, it is not clear that the metric adequately compensates for the generally lower levels of academic preparation, need for developmental coursework, and older ages that tend to be associated with greater work and family obligations among community college students.

Distinct standards for four-year institutions and programs and subbaccalaureate programs can alleviate this problem. Even if it is feasible to adjust earnings metrics to estimate the return on investment accounting for typical time to degree and tuition payments, it is likely unreasonable to set one standard for both bachelor's degrees and short-term certificate programs. Earnings paths are different for different types of occupations, so the choice of a point in time for measurement will have a different impact on different types of institutions and programs (Carnevale, Rose, and Cheah 2014; Deming and Noray 2020).

A natural corollary is to consider program-level, rather than institution-level, accountability (Blagg et al. 2021). Graduate programs, for example, have recently come under scrutiny for high debt-to-income ratios, an outcome resulting from the lack of borrowing limits for graduate loans. At minimum, graduate program accountability within an institution should be separate from undergraduate program accountability, and the variation in graduate programs (compare an MA in education with an MD) makes treating them separately appealing. Similarly, it may not make sense to compare cosmetology programs with medical billing programs. At the same time, it is likely reasonable to keep most bachelor's degree programs at many colleges grouped together, unless there are separate colleges within the university dedicated, for example, to engineering and nursing.

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, which span academic outcomes, student debt outcomes, and labor market outcomes, they will be able to avoid some of the vulnerability associated with particular students and program characteristics.

# Principles

Before we examine accountability metrics proposed by others and investigate the potential impact of alternative metrics, it is helpful to develop principles for an accountability system and a framework for evaluating different approaches.

## Use Multiple Metrics and Allow Flexibility for Meeting Standards

*Instead of excluding programs and institutions just because of, for example, high default rates, allow them to meet, say, three out of four standards in the areas of loan repayment, completion, and employment outcomes.*

Accountability standards aim to hold programs and institutions responsible for providing high-quality, effective educational opportunities. Measuring success is not easy, and any metric will be an imperfect proxy for the desired outcomes.

The empirical results discussed below illustrate that reasonable metrics yield different results. A program may have a high completion rate but poor employment outcomes. Students may accrue high levels of debt but succeed in repaying that debt. Students may default on their loans at high rates despite finding employment in their field of study.

Choosing only one metric will thus advantage some programs and disadvantage others of arguably comparable quality. Moreover, it is frequently possible to game the system by manipulating measurements. For example, to avoid sanctions resulting from high default rates, some institutions push students into forbearance on their loans even when that is not a good choice for them (GAO 2018). Requiring that programs meet multiple metrics will make this sort of manipulation more difficult.

The current accountability system—under which institutions that fail face serious all-or-nothing penalties if they have an extremely poor showing on one metric (loan default) for multiple years—has sanctioned only a handful of institutions. High-stakes, all-or-nothing standards are likely both to encourage institutions to find workarounds and to inappropriately place some programs on the wrong side of the line with serious repercussions. Both the use of multiple metrics and allowance for graduated consequences rather than all-or-nothing accountability should mitigate these problems. Providing warnings and time for improvement will permit programs with the motivation and potential for strengthening outcomes to do so.

## Penalties Should Not Be All or Nothing

*Instead of saying there is no problem if the default rate is below, for example, 25 percent but excluding the school from federal student aid programs if the default rate hits 25 percent, regulations should include a zone for added scrutiny or probation.*

An all-or-nothing threshold increases incentives to manipulate metrics and diminishes the opportunity for programs and institutions to improve outcomes. Sanctions in the warning zone should be designed to increase reporting, allow for monitoring, and provide information about best practices and other guidance to make it possible for enrolled students to receive meaningful opportunities.

Sanctions should not involve onerous payments that will exacerbate the financial problems of institutions with inadequate resources and make it more difficult for institutions to improve outcomes. Reducing the amount of financial aid available, limiting access to either grants or loans, fining institutions, or imposing other financial penalties can only reduce student success. Instead, institutions in the "warning zone" might be subject to additional reporting requirements, be expected to implement and evaluate new student support systems, or modify other practices associated with poor student outcomes and then face exclusion from federal student aid programs if they fail to improve.

## Standards Should Consider the Distribution of Student Outcomes, Not Just the Average Outcome

*Standards could, for example, focus on the share of students earning above the average high school graduate (or some other minimum threshold), not just average earnings relative to the threshold.*

Averages can hide considerable variation in student outcomes. In one program with average debt of $20,000, 90 percent of students may owe between $15,000 and $25,000. In another program with the same average debt, 20 percent of students may have borrowed more than $50,000, with a similar share of students having accrued low levels of debt. Similarly, a few stars in an acting program might generate average cohort earnings unrepresentative of outcomes for most students.

Particularly with earnings, the goal should not be to maximize average earnings, discouraging the study of education and entry into public service occupations. Instead, it should be to ensure that as many students as possible end up able to earn a living wage.

## Where Appropriate, Standards Should Include Students Who Enroll but Do Not Graduate in Addition to Graduates

*Measuring debt repayment rates for all entering students, not just those who complete credentials, provides a more thorough picture, particularly for institutions with low completion rates.*

Students who do not complete their programs are less likely to successfully repay their loans and are more likely to default than those who graduate. Ignoring these students gives an incomplete picture of a program's outcomes. It may not be appropriate to include noncompleters when evaluating earnings, but that should not exclude them from the loan metrics.

## Incorporate Differences in Student Bodies in Making Judgements, but Set Minimum Thresholds for All Students to Avoid Institutionalizing Weak Expectations for Students Facing the Greatest Barriers to Success

*It is unrealistic to expect institutions enrolling large shares of students with weak academic preparation or challenging life circumstances to have the same outcomes as selective institutions. But all students deserve to benefit from meaningful standards.*

Completion rates are highly correlated with the characteristics of incoming students. Among first-time full-time bachelor's degree–seeking students first enrolling in 2013 at four-year institutions, six-year completion rates ranged from 29 percent for open-admissions institutions and 52 percent at institutions accepting at least 90 percent of applicants to 74 percent at those accepting 25.0 to 49.9 percent of applicants and 89 percent at those accepting less than 25 percent of applicants.[1] Postcollege earnings are also correlated with student characteristics (Klor de Alva 2022).

Any standard that applies to all institutions will have to be low enough to allow institutions most successful at educating students facing the greatest barriers to success to pass and will be unlikely to create any incentives or lead to any improvement at selective institutions. That may be an acceptable outcome if the primary goal of the accountability system is to weed out the worst performers.

The alternative is to find a reasonable way of setting standards that are higher for schools with better-prepared students from higher socioeconomic backgrounds. But with this approach, it is critical to set minimum thresholds for all institutions and programs. No matter how many challenges students face, programs that improve life for almost none of them are unacceptable.

Adjusting earnings for demographics is different from adjusting completion rates. A salient issue is that labor market discrimination, combined with other factors, generates earnings differences by race, ethnicity, and gender, despite similar levels of education and qualifications. Sanctioning institutions that educate larger shares of Black students or larger shares of women is inappropriate. Optimally, postcollege earnings would be compared with precollege earnings for adults in the same geographic location with similar demographic characteristics (including socioeconomic status). A potential (partial) solution is to adjust for earnings in zip codes of student origin when measuring the returns on investment and possibly for defined demographic groups when measuring earnings against a current threshold.

## Set Standards That Apply to Institutions in All Sectors, but Recognize the Particular Risks of the For-Profit Structure and Limited Resources of Some Institutions, Particularly Community Colleges and Minority-Serving Institutions

*All institutions should be held accountable for student outcomes. But this reality does not conflict with regulations that focus on financing structures, which differ considerably across sectors and governance structures.*

By definition, for-profit institutions have goals and incentives different from those of other institutions. These institutions are most dependent on tuition revenues and get the largest share of revenues from federal student aid. Public colleges and universities are subject to state oversight. Public and private nonprofit institutions are responsible to boards that have no financial interest in the institutions. The absence of these safeguards in the for-profit sector and the role of the profit motive make it appropriate for the federal government to apply extra scrutiny.

## Distinguish between Four-Year, Two-Year, and Less-Than-Two-Year Programs and Institutions When Setting Standards

*Neither a single earnings level nor a single time frame is adequate for measuring success in programs of all lengths.*

Absent a reliable measure of the return on investment in higher education for individual students, the threshold for acceptable earnings should be lower for short-term programs than for bachelor's degrees and should be even higher for advanced degree programs. One option would be to compare program graduates with others with similar credentials.

The value and feasibility of incorporating completion rates for part-time or transfer students will differ by type of institution, as will expectations about a minimum acceptable completion rate.

## Design Standards That Account for Geographic and Cyclical Variation

*Because of earnings differences across demographic groups and geographic locations, using one national earnings threshold to determine whether program graduates get a reasonable return on their investment would be unreliable.*

It is possible to adjust earnings of program graduates for earnings levels where they live at time of college enrollment. Although postcollege earnings depend on where students live after they leave school, basing adjustments on precollege location provides a better measure of the earnings increment resulting from postsecondary education.

To avoid a disproportionate number of failures when the labor market is weak and passing marks when it 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.

## Balance the Importance of Measuring Outcomes Allowing a Reasonable Amount of Time after Leaving School with Ensuring That Problems Do Not Persist for a Long Time with No Accountability

*Earnings immediately after college do not necessarily reflect long-term career paths. But outcomes of students who enrolled a decade ago may not reflect current students' experiences.*

Minaya and Scott-Clayton (2020) find that earnings measured in the first couple of years after program completion are poor indicators of long-term outcomes. Moreover, earnings paths differ among fields, so some graduates likely take longer than others to establish themselves in the labor market (Speer and Leighton 2020).

## Metrics Should Be Simple, and Data Should Be Available

*Metrics should be simple enough for institutions (and possibly students) to understand their meaning and respond to them.*

Attempts to make metrics sensitive to the characteristics of incoming students by, for example, using regression analysis to determine completion probabilities for individual students might prevent

institutions from being penalized for enrolling students facing the greatest barriers to success. But in addition to the problems with setting low standards for some students, it will generate metrics that are difficult for institutions to predict and understand.

It might be possible to design metrics that would be reliable if accurate data were available but that lose their advantages with the limitations of existing available data. For example, the return on investment measures the value of a postsecondary credential more effectively than a simple earnings threshold. But the complexity of estimating counterfactual earnings, as well as net prices and time enrolled, for students with different outcomes may produce misleading results.

Pointing to additional metrics for which data could reasonably become available is constructive, but developing a set of imperfect but reasonable metrics is more sensible than abandoning the effort in the absence of perfection.

## Setting Benchmarks

Two closely related categories of proposals for benchmarking are those that focus only on loans (frequently through "risk sharing" or holding institutions accountable for a portion of the loans their students do not repay) and broader accountability proposals. The risk-sharing proposals, designed to provide incentives for institutions to improve outcomes, rely on default or loan repayment rates. These systems are generally not pass-fail but provide a method for developing financial penalties for institutions that correspond to their student debt performance.

Broader accountability proposals are generally based on a range of student outcomes, which may incorporate labor market outcomes or completion rates, in addition to or instead of debt repayment. We argue that accountability measures designed to prevent students from taking their federal financial aid to programs and institutions where they have little chance of emerging with valuable credentials will be more equitable and reliable if they are based on multiple metrics measuring a range of outcomes.

Some current regulations focus on *institutional metrics* rather than student outcomes. For-profit institutions cannot obtain more than 90 percent of their tuition revenues from federal student aid. The financial responsibility composite score monitors the financial positions of private nonprofit and for-profit institutions. In addition, the incentive compensation ban prohibits institutions from paying employees based on the number of students they recruit or who receive financial aid. (See appendix A for an overview of existing regulations.)

Our focus is on student outcomes, and below, we discuss a set of possible metrics, first individually and then as integrated accountability standards. Data from the US Department of Education's College Scorecard allow us to explore several potential accountability metrics. Despite the strong arguments for applying program-level benchmarks where possible, the discussion below focuses on institutions. Our analysis is illustrative, and the institution-level examples are more manageable (and data are more available).

The data include 5,346 postsecondary institutions (table 1). Almost half of these are four-year institutions, which enroll 68 percent of undergraduate students. Two-year institutions account for just under one-third of institutions and a similar share of undergraduates. Because less-than-two-year institutions tend to be small, they account for about one-quarter of all the institutions but enroll only about 1 percent of students. In other words, non-degree-granting institutions could account for a significant share of those not meeting accountability thresholds, but a very small share of students would be affected by this outcome.

TABLE 1

**Institutions and Undergraduate Students, by Type and Sector**

| | Institutions (5,346) | Students (15,236,373) |
|---|---|---|
| **Four-year** | | |
| Public | 13% | 46% |
| Private nonprofit | 29% | 18% |
| For-profit | 4% | 4% |
| All four-year | 46% | 68% |
| **Two-year** | | |
| Public | 18% | 29% |
| Private nonprofit | 2% | 0% |
| For-profit | 8% | 1% |
| All two-year | 28% | 30% |
| **Less than two-year** | | |
| Public | 4% | 0% |
| Private nonprofit | 1% | 0% |
| For-profit | 21% | 1% |
| All less than two-year | 26% | 1% |

**Source:** Authors' calculations using College Scorecard data.
**Note:** According to the Integrated Postsecondary Education Data System, two-thirds of four-year institutions account for two-thirds of degree-granting institutions and a similar share of undergraduate students. Less-than-two-year institutions are not included in these data because they award only certificates, not associate or bachelor's degrees (Digest of Education Statistics 2020, tables 317.10 and 303.50).

Using the College Scorecard data, we look at metrics for default, loan repayment, completion, and earnings. These are not the only possible areas for developing accountability metrics, not even the best ones. Unusually rapid enrollment growth or decline might, for example, be a reliable warning sign.[2]

Devising a reliable and functional accountability system for postsecondary institutions will inevitably involve many judgments and compromises. There would not be a perfect set of metrics even if high-quality nuanced data were easily available. A fundamental issue is that measures such as those included here capture only a portion of what we expect higher education to deliver. Student learning, for example, is not on the list of potential metrics because it is so difficult to define and measure.

Nonetheless, it is critical we develop a system that does the best possible job of directing both students and federal financial aid dollars toward institutions and programs that will serve them well.

The principles we have laid out do not lead to one unique set of metrics. The discussion below focuses on the trade-offs involved in selecting specific metrics and uses examples of reasonable metrics to illustrate potential outcomes for students and institutions.

To anchor the discussion, we use the example of thresholds set at levels where 5 percent of students would attend institutions failing the standard. 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. Otherwise, institutions would be chasing a moving target, and a similar share of students would attend failing institutions whether overall outcomes improved or worsened. Rather than choosing one percentile threshold for all metrics, it may be preferable to modify the threshold for each metric to avoid as much as possible choosing points in steep sections of the distribution, where a small change in the cutoff has a large impact on the number of institutions on either side of the line.

## Successful Loan Repayment

*Share of borrowers or share of dollars? Share of borrowers or share of students?*

Both default rates and repayment rates are indicators of manageable debt levels. The current regulatory system relies on cohort default rates (CDRs), measuring the share of borrowers entering repayment who default in a specified period. But the system could instead focus on the share of borrowers not reducing their loan balances, the share of all students (whether they borrowed or not) not managing to repay their loans, or the share of dollars borrowed being repaid.

Focusing on students gives an indication of the share of students for whom the 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.

Ignoring how much debt students incur and whether their failure to repay is placing a significant burden on taxpayers, as the CDR does, is also questionable. Are borrowers who have paid down $1 of their $5,000 debts after five years in the same situation as those who have paid down $1 of their $40,000 debts? Is a school where 15 percent of borrowers have defaulted on small loans comparable with a school where 15 percent of borrowers with high debt levels have defaulted?

We address these trade-offs below by using the CDR, which is based on borrowers, and the share of debt outstanding, which is based on dollars. We weight the CDR by share of students with debt to ensure that institutions where few students borrow score better than those where most students borrow but do not adjust the dollar-based repayment rate for share of students borrowing. It is more problematic to ignore the number of students borrowing for a metric based on the number of students with poor outcomes (default rate) than for a metric based on overall dollars borrowed.

## Default Rates

Table 2 illustrates the impact of incorporating the share of students with debt into the default rate. The average institutional default rate in the College Scorecard data (weighted by the number of borrowers at each institution) is 8 percent. Multiplying by the share of students borrowing to focus on the share of students—rather than the share of borrowers—who default lowers the average to 4 percent, indicating that only about 50 percent of undergraduate students have federal student loans.

Because a small share of students at public two-year colleges borrows, using the CDR adjusted by the share of students borrowing reduces the average CDR from 16 percent to 4 percent for this sector and moves public two-year colleges from the highest to among the lowest default rates.

Adjusting for the share of students borrowing is critical to appropriate treatment of public two-year colleges.

TABLE 2

**Unadjusted and Adjusted Average Cohort Default Rates**

| | Average CDR (weighted by number of borrowers) | Average adjusted CDR (weighted by number of borrowers) |
|---|---|---|
| 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% |
| **Total** | **8%** | **4%** |

**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 adjusted CDR for each institution is calculated 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.

If the CDR is to be a part of an accountability system, where and how should the threshold be set? Currently, the threshold for Title IV eligibility is 40 percent in a single year or 30 percent in three consecutive years. Only a handful of schools have failed the current test.

Proposals for new benchmarks related to default rates include one from Baum and Schwartz (2018), who suggest using more than one default metric to account for both the share of borrowers who default and the share of the cohort's loan dollars that go into default. The analysis below includes a default rate based on students and adds a repayment rate based on dollars.

Ahlman, Cochrane, and Thompson (2016) focus on students rather than balances, suggesting that the default rate be combined with a repayment rate for accountability purposes. Their Student Default Risk Indicator equals the share of students borrowing times the CDR and measures the probability that a student will experience default. The authors propose that schools would lose access to Title IV funds if at least 20 percent of students have defaulted on their loans. This could involve, for example, a 40 percent default rate among the 50 percent of students who borrow or a 100 percent default rate among the 20 percent of students with debt.

In his proposal for institutional risk sharing for student loans, Hillman (2016) also argues for using both a default metric and a repayment metric, with harsher penalties for those that fail both tests. He proposes setting benchmarks in terms of standard deviations from the current mean. Hillman agrees

that the debt measures should be weighted by the share of students borrowing, so even those with low shares of borrowers would be held accountable but their thresholds would be higher.

An alternative to weighting by the share of students who borrow is to exclude institutions at which a low share of students borrow from the system of sanctions. Klor de Alva and Schneider (2016) suggest that only institutions with at least 25 percent of students participating in the federal student loan program would be subject to risk-sharing provisions, which would require payment of a share of the loans in default. One problem with this approach is that setting such a cutoff creates a cliff, where an additional student borrowing could lead to a dramatically different outcome for the institution.

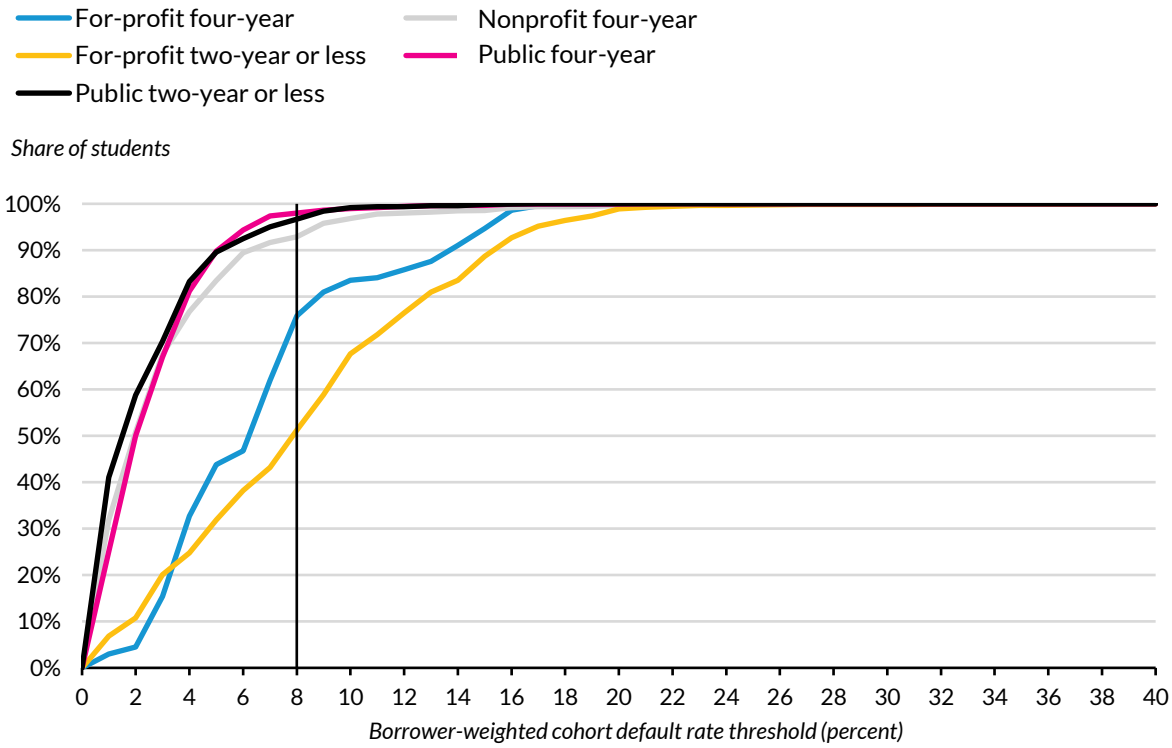
To more accurately target protection for students, it would be preferable to rely on default rates for individual programs where appropriate. Even without this detailed breakdown, the metrics could be calculated separately for undergraduate and graduate students. Aside from combining these two distinct groups—which has a very different impact on different types of institutions—the available default rates exclude PLUS loans, both those issued to parents of undergraduate students and those issued to graduate students (FSA 2020). Correcting this omission would make the default rates a more accurate measure of student loan repayment.

Figure 1 shows the share of students in each sector attending institutions that would pass the adjusted CDR metric at each potential threshold. If the threshold were an adjusted default rate of just over 8 percent, institutions serving roughly 5 percent of students would fail. At this threshold, 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 3)—reflecting the reality that few community college students borrow. (In the unadjusted CDR [appendix figure C.1], community colleges fare far worse, with 63 percent of students at failing institutions attending public two-year schools.) For-profit institutions 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 3

**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% |
| **Total** | **810,416** | **16,167,992** | **5%** | **755** | **4,785** | **16%** | **8%** |

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

## Repayment Rates

The standard repayment rate defined in the College Scorecard data is the share of borrowers in a cohort who reduce their loan principal by at least one dollar in a specified number of years. The College Scorecard reported repayment rates at one, three, five, and seven years after entering repayment until 2016–17. Currently, the College Scorecard reports the share of students in each of eight possible repayment statuses one and two years after entering repayment: in default, delinquent, in forbearance, in deferment, in good standing but not making progress, making progress, paid in full, and discharged. These breakdowns are also available by field of study for completers.

A different repayment metric is the share of dollars repaid after a certain number of years. Recently, the College Scorecard has begun reporting the share of dollars outstanding after 1, 4, 5, 10, and 20 years. These measures are available separately for undergraduate, graduate, and Parent PLUS loans, enabling separate accountability rules for undergraduate and graduate programs. In either case, policymakers face the trade-off of relevancy for current students (a more recent measure) versus

reliability (a longer-term metric). In general, we find that very recent measures do not distinguish well among institutions (see appendix B) and recommend the use of about 5 years.

Focusing on borrower-based repayment rates, Baum and Schwartz (2018) looked at how many students would be able to pay down some principal on the average debt under income-driven repayment, which generally requires payments equal to 10 percent of the borrower's income exceeding 150 percent of the federal poverty level. Based on the share of young associate degree holders with earnings high enough to generate payments that would reduce typical loan balances, Baum and Schwartz concluded that at least 40 percent of former students should be repaying at least some of their loan principal within three years.

The loan repayment rate Ahlman, Cochrane, and Thompson (2016) propose is a Student Non-Repayment Risk Indicator, which parallels their default metric. It equals the share of students borrowing times the nonrepayment rate (the share of borrowers failing to reduce their loan principal by at least one dollar).

In contrast, Webber (2018), focusing on loan dollars, proposes penalties based on the student loan repayment rates of former students and bonuses based on the performance of student groups facing barriers to success such as Pell grant recipients. The size of the penalty would be based on the level of balances not being reduced two years into repayment.

### DOLLAR-BASED REPAYMENT RATES

An alternative to measuring the share of students successfully repaying their debts is to focus on the share of dollars repaid.

Chou, Looney, and Watson (2017) propose requiring schools to pay a portion of the outstanding debt if borrowers have not repaid at least 20 percent of the amount borrowed after 5 years. For accountability purposes, the implication is that the share of principal repaid after 5 years would be a useful metric. Under a standard 10-year repayment plan with a 6.8 percent interest rate, 40 percent of the principal would be repaid after 5 years. The 20 percent repayment threshold would put borrowers on target to fully repay within about 15 years. Many institutions that do not have high default rates have low repayment rates, indicating that their students do not earn enough to repay their loans. This repayment rate is difficult to game or manipulate. Under this standard, about half of institutions would pay some penalty, which would be a severe outcome for an accountability threshold. Lengthening the projected repayment term from 15 years to 20 years would lead to a standard of 14 percent of principal repaid after 5 years.

Using a metric related to the share of balances repaid to complement the 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. A repayment rate based on dollars provides different information from the default rate based on numbers of students.

Dollar-based repayment rates reflect the share of loan dollars outstanding, relative to the amount originally borrowed. Based on a standard 10-year repayment plan and a 4.5 percent interest rate, roughly 56 percent of the original loan balance should be outstanding after 5 years. But this is rarely the case in practice (table 4; figure 2). Rather, at the average institution, it appears that undergraduate students in aggregate have not made much progress on their loans, and many of them are negatively amortized, with unpaid accrued interest increasing their outstanding balance beyond the original amount borrowed.

TABLE 4

**Average Share of Undergraduate Loan Dollars Outstanding after Five Years**

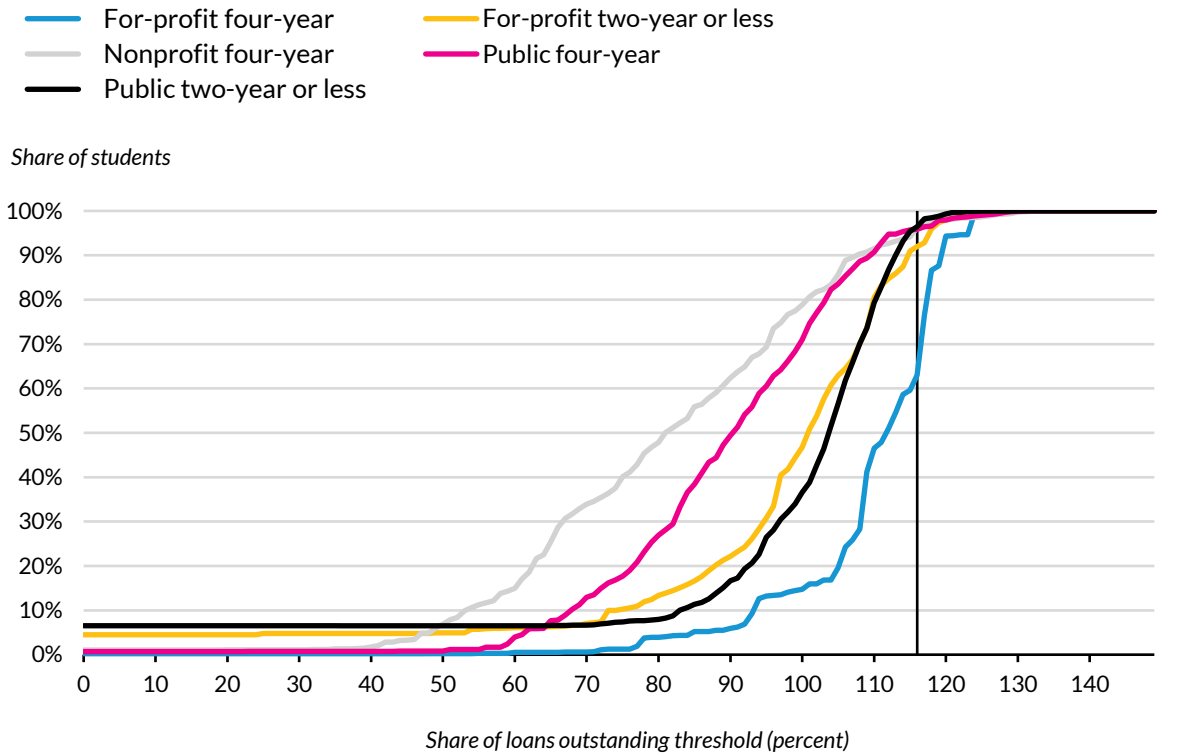
| | Average | Average (weighted by number of borrowers) |
|---|---|---|
| For-profit two-year | 97% | 103% |
| For-profit four-year | 103% | 111% |
| For-profit less-than-two-year | 83% | 99% |
| Nonprofit two-year | 78% | 103% |
| Nonprofit four-year | 80% | 87% |
| Nonprofit less-than-two-year | 65% | 95% |
| Public two-year | 83% | 105% |
| Public four-year | 92% | 91% |
| Public less-than-two-year | 56% | 96% |
| **Total** | **84%** | **94%** |

**Source:** Authors' calculations using 2018–19 College Scorecard data.
**Note:** Because of data limitations, sector averages weight institutions by number of borrows rather than dollars of debt.

FIGURE 2

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

*Share of dollars outstanding*




URBAN INSTITUTE

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

The 95th percentile of the share of loan dollars outstanding is a startling 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 the repayment period). One-third of undergraduate students in the for-profit four-year sector attend institutions where the outstanding loan balance after five years is more than 117 percent of the amount borrowed (table 5). The shares of students attending such institutions in other sectors are low, but 42 percent of undergraduate students overall attend institutions where the share of dollars outstanding is greater than the initial balance after five years. (Appendix figure B.2 shows the share of loans outstanding among graduate student borrowers.)

TABLE 5

**Failing the Five-Year Repayment Rate Threshold**

*Undergraduate students and institutions*

| | Students attending failing institutions | Total students | Share | Failing institutions | Total institutions | Share | Failure threshold |
|---|---|---|---|---|---|---|---|
| For-profit two-year | 29,722 | 264,857 | 11% | 43 | 399 | 11% | 117% |
| For-profit four-year | 256,603 | 787,593 | 33% | 37 | 174 | 21% | 117% |
| For-profit less-than-two-year | 7,001 | 199,588 | 4% | 45 | 866 | 5% | 117% |
| Nonprofit two-year | 1,797 | 43,798 | 4% | 8 | 112 | 7% | 117% |
| Nonprofit four-year | 125,972 | 3,603,996 | 3% | 93 | 1,301 | 7% | 117% |
| Nonprofit less-than-two-year | 150 | 6,614 | 2% | 1 | 32 | 3% | 117% |
| Public two-year | 110,055 | 3,516,512 | 3% | 22 | 917 | 2% | 117% |
| Public four-year | 291,154 | 7,600,415 | 4% | 45 | 680 | 7% | 117% |
| Public less-than-two-year | - | 42,522 | 0% | 0 | 215 | 0% | 117% |
| **Total** | **822,454** | **16,065,895** | **5%** | **294** | **4,696** | **6%** | **117%** |

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

## Completion Rates

Including completion rates 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. There are some questions about whether 150 percent of normal time (three years for an associate degree and six years for a bachelor's degree) is long enough to capture the full picture, but there is no need to wait for students to settle into the labor market or establish their repayment patterns.

Official IPEDS completion rates are flawed because they include only first-time full-time students. Completion rates are available for part-time students, but only a small share of institutions report this outcome. Data from the National Student Loan Data System include all enrolled students receiving Title IV aid but mix full-time and part-time students. Comparing the share of students completing within a specified number of years without distinguishing between these groups would disadvantage institutions enrolling large shares of part-time students, whose completion times will inevitably be longer. Again, there is no perfect choice.

Accordingly, we focus on the share of first-time full-time students completing their programs within 150 percent of normal time. A few institutions have higher completion rates for part-time students than for full-time students, and some accommodation for this circumstance might be made in the accountability system. In appendix D, we explore the rates that include all Title IV students to get some insight into the impact of including students who transfer and earn credentials at institutions other than the one in which they first enrolled. Appendix figure D.4 plots five completion rate metrics to show how they compare.

Community colleges perform poorly on the IPEDS completion metric, but in contrast to the debt metrics, some for-profit institutions (particularly two-year and less-than-two-year institutions) tend to have relatively high completion rates. Arguably, the missions of community colleges differ substantially from other institutions: they are designed as low-cost, open-access institutions. Allowing students to explore their interests and skills without a large financial investment is not necessarily a bad thing. It may be less of a problem for students to spend some time at a community college without completing a credential than to enroll in a more expensive institution or one with the sole mission of preparing students for specific occupations.

The overall enrollment-weighted average completion rate (150 percent of normal time) for all institutions is 51 percent, ranging from 27 percent for public two-year colleges and 35 percent for for-profit four-year institutions to 71 percent for nonprofit less-than-two-year institutions. Again, students face fewer barriers to completing very short–term programs. The non-degree-granting institutions, which offer only short-term programs, have the highest completion rates in all three major sectors—public, private nonprofit, and for-profit. Four-year for-profit institutions have lower completion rates than shorter-term for-profit institutions, but two-year institutions have the lowest completion rates in the public and private nonprofit sectors.

TABLE 6

**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% |
| **All institutions** | **51%** | | | | | | |

**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.

There is more variation in completion rates within some sectors than in others. In the small private nonprofit two-year sector, the 25th percentile completion rate is 36 percent and the 75th percentile is 83 percent. In contrast, at public two-year colleges, this range is from 22 percent (25th percentile) to 39 percent (75th percentile).

Baum and Schwartz (2018) propose that at four-year institutions, the threshold for eight-year completion rates should be the current average official on-time completion rate of 40 percent. At two-year institutions, the threshold for the completion rate within 200 percent of normal time should be 25 percent—approximately the current average completion rate within 150 percent of normal time for public two-year colleges. Relying on completion within 150 percent of normal time rather than 200 percent would call for lower thresholds.

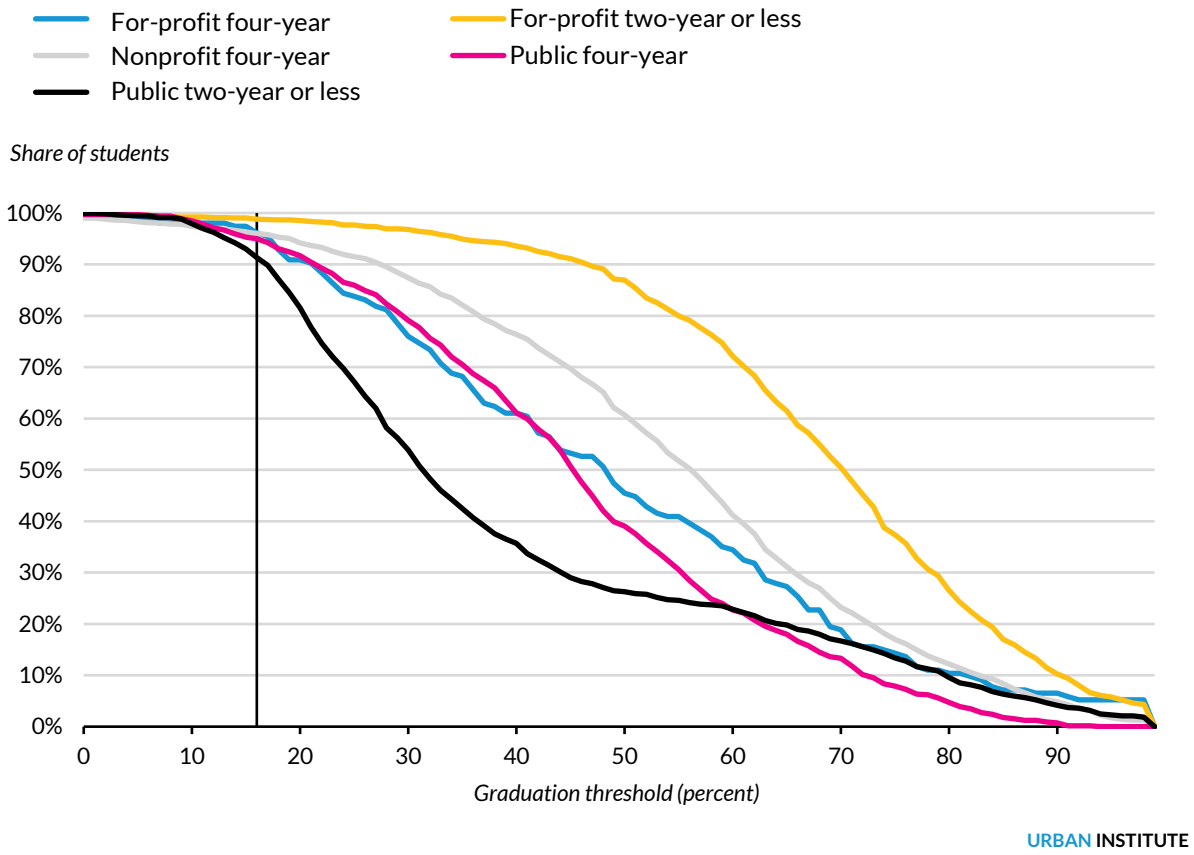
Klor de Alva and Schneider (2016) propose a risk-sharing system that would be based in part on CDRs and in part on the graduation rates of Pell grant recipients (three-year graduation rates for two-year schools and six-year graduation rates for four-year schools).

Blagg and Chingos (2016) argue that metrics based on completion rates (in addition to postcollege earnings) are preferable to those based on student loan performance because they do not require a long time frame to assess outcomes and because completion is a strong predictor of successful loan repayment.

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).

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.

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 quite different shares of students at different types of institutions. Setting thresholds separately, so that 95 percent of students in each of the three institution categories 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. Among four-year

institutions, for-profit institutions are least likely to pass the threshold, but among two-year and less-than-two-year institutions, private nonprofits perform worst (appendix figures D.1 through D.3).

## 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. Postcollege earnings are a reasonable metric for institutional accountability. Another possibility would be to measure employment or unemployment rates.

The College Scorecard data provide measures of earnings along with corresponding data on employment 6 years, 8 years, and 10 years after enrollment. They include information for subpopulations within institutions.

### ALL STUDENTS OR ONLY GRADUATES?

It is logical to include all students who enroll in evaluating the experiences of students with college debt. But should institutions be held accountable for the earnings of students who leave without completing their programs? One potential argument in favor is that institutions should strive to improve students' lives and not simply enroll students with little chance of success. But including a graduation rate metric helps offset this concern.

### WHEN SHOULD EARNINGS BE MEASURED?

Earnings in the first year after graduation are unlikely to be an accurate representation of the long-term payoff of a college degree. It takes some students a while to find their place in the labor market. Some fields require apprenticeships or internships. Earnings trajectories vary across occupations. But the longer the time frame for measuring earnings, the longer the wait for failing institutions to be penalized and the less likely the outcome being measured will be the same for current students. About five years after completion is a reasonable compromise for measuring earnings.

### GEOGRAPHIC DIFFERENCES

If graduates tend to enter the labor market in the geographic area where they studied, earnings levels for students from schools in urban areas are likely to be higher than those for students from similar schools in rural areas. Workers in New York earn more than workers with similar skills in Wyoming. Cost-of-living differences are a real challenge, but it should be possible to compare earnings of graduates with the earnings of others in the same geographic area, adjusting benchmarks accordingly.

### ADJUSTING FOR PROGRAM TYPE

Earnings of typical bachelor's degree recipients are higher than earnings of associate degree recipients, which are higher than earnings of adults holding short-term certificates. Setting one standard, such as the share of graduates earning more than 150 percent of the federal poverty level or more than the earnings of the average high school graduate (or relying on average earnings) will unduly disadvantage shorter-term programs. These programs should also boost earnings, but there should be a different standard for judging success.

### AVERAGES AND DISTRIBUTIONS

Proposed accountability standards frequently rely on averages. How does the average debt level compare with average earnings? Do average earnings exceed a specified threshold? But averages conceal important differences across programs and institutions.

Perhaps most significantly, maximizing average earnings of graduates is a questionable goal. Should institutions be pushing their students to become investment bankers instead of teachers and social workers? Should students be discouraged from pursuing artistic ambitions? If the goal is to ensure most students can support themselves at a reasonable standard of living, it is more reasonable to use a metric based on most students (60 percent, 75 percent, or 90 percent?) earning above some threshold.

### EARNINGS VERSUS RATE OF RETURN

We have framed the difficulties in measuring earnings in terms of earnings levels. But the measure of the success of an educational experience is the earnings increment it generates, not the absolute level of graduates' earnings. Bachelor's degree recipients should not only earn more than associate degree recipients. The earnings increment beyond what graduates would have earned after high school should be larger. Four years (or five or six years) of paying tuition and being at least partly out of the labor market is a bigger investment than two years. A similar rate of return requires a larger absolute return.

But measuring rate of return is more challenging than measuring the other outcomes we have discussed. Attempts to do this include Itzkowitz, which looked at average net tuition costs and normal time to degree and compared earnings 10 years after enrollment net of estimated costs with the earnings of typical high school graduates in the state.[3]

It may be more feasible and meaningful to compare earnings outcomes with the earnings of other adults with similar levels of education. What share of the two-year institution's graduates earn more than the 25th percentile for all adults (probably within an age group) with associate degrees? This

metric would not be a measure of value added (comparing earnings with and without the degree) but would be a measure of whether the degrees measure up to similar degrees.

Briones and Turner (2021) probed the College Scorecard data on earnings to further understand the reliability of available data, particularly to assess outcomes for older students. They find that missing data is a problem for private nonprofit and for-profit two-year and less-than-two-year institutions but that when earnings data are available, there is a strong correlation across cohorts and among different subgroups. Students at four-year institutions and community colleges experienced earnings growth of more than 20 percent between the 6- and 10-year points of observation; earnings growth for the less-than-two-year sector was much weaker for the institutions for which data were available.

The authors argue that taking local labor markets into account is critical for evaluating earnings data. Differences by program are also stark, raising concerns about the absence of program-level data for small, short-term institutions in the College Scorecard data. Minaya and Scott-Clayton (2020) use state administrative data from Ohio to examine consistency across several labor market metrics and a measure of program completion. They find that controlling for student characteristics matters for measures immediately after college but is less significant if labor market outcomes are measured a few years out, which makes those measures more reliable indicators. The authors find that labor market outcomes add considerable information to graduation rates but that the choice of outcome measures (e.g., earnings, employment stability) can make a big difference. They also find that the patterns are different for four-year and subbaccalaureate institutions.

Baum and Schwartz (2018) propose a benchmark based on the distribution of income among adults with different levels of educational attainment. In 2015, about 60 percent of associate degree holders with earnings made more than the median for high school graduates with earnings. Accordingly, the authors propose a benchmark requiring that 60 percent of an institution's enrolling students should earn more than the median for high school graduates within 10 years of enrolling.

Blagg and Chingos (2016) would penalize institutions where a large proportion of students earn less than a specified minimum (e.g., the federal poverty level) for multiple years.

Because of the imperfect correlation between potential metrics, it is important to combine labor market and academic outcomes. Using multiple metrics is the only way to develop a reliable accountability system, given the inconsistencies across metrics and the range of circumstances individual institutions face. Moreover, it is appropriate to evaluate four-year institutions separately from those that offer primarily shorter programs.

### SHARE OF STUDENTS EARNING ABOVE A THRESHOLD

Although not all the return on higher education is (or should be) financial, most students should be financially better off than had they not attended college. Of course, we do not see a student's counterfactual outcome—their earnings had they not attended college. But how do we set a reasonable threshold? The College Scorecard data offer a few options, and we suggest a few additional possibilities.

***Share of graduates earning above a threshold.*** The College Scorecard reports the share of completers earning at least 150 percent of the federal poverty level for a single person (about $19,000 in 2021) three years after graduation. Although this seems like a low bar, only in the public and private nonprofit four-year sectors do most institutions have more than 80 percent of students reach this threshold—highlighting the problem with using one standard for all types of programs and institutions and the disturbingly low earnings of many former students.[4] A metric based on earnings above 150 percent of the federal poverty level will not differentiate among four-year institutions. Using a threshold of the earnings of typical high school graduates will face a similar problem.

A possible alternative would be a metric based on the share of graduates who receive Supplemental Nutrition Assistance Program or other benefits. The College Scorecard does not collect this information, and it would require state-federal data-sharing agreements, but such a partnership is worth exploring. Another tweak policymakers could consider is using household income (and poverty levels) rather than individual income, as some individuals may have low or zero earnings by choice but have adequate household income. This approach, however, would break the direct link between schooling and earnings.

Any metric comparing the earnings of graduates with an income threshold should either use a different threshold for two- and four-year programs or incorporate a different expectation about the share of graduates whose earnings exceed the threshold.

***Comparing earnings for an institution's students with the earnings of students at similar institutions.*** Rather than comparing earnings with the federal poverty level, we can look at the 25th percentile of earnings, which allows the threshold to be set separately for four-year, two-year, and less-than-two-year institutions. Earnings are relative to others from the same type of institutions, rather than an external benchmark. These data are reported for all borrowers, not just completers, and are several years out of date. We recommend that earnings metrics be reported for completers only, so the values we show here will appear low. 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 is $16,700 (figure 4). In other words, 5

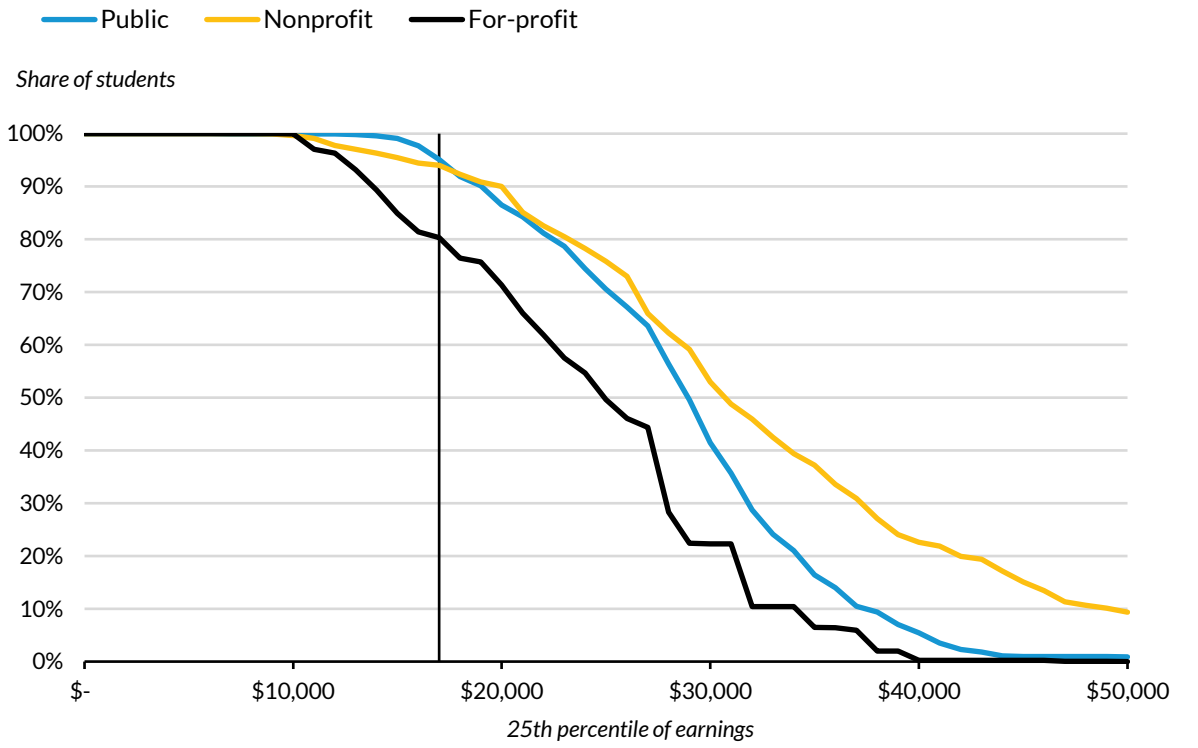
percent of students attend institutions where more than 25 percent of students earned less than $16,700 10 years after enrolling. For-profit institutions are less likely than public and private nonprofit four-year institutions to pass this metric.

Appendix figures E.1 and E.2 show similar data for two-year and less-than-two-year institutions. The relative performance of public, private nonprofit, and for-profit institutions depends on where the threshold is set.

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.

***Do at least three-quarters of an institution's students earn more than a threshold specific to the program length and adjusted for median household earnings in the zip codes the students came from?*** In addition to using different thresholds for institutions offering different types of credentials, it is possible to adjust for geographic differences in earning levels. Consider community college students in rural Virginia compared with those in Arlington, a wealthy DC suburb. Those from Arlington are likely to earn more

regardless of whether Northern Virginia Community College (mean earnings are $47,200) is in fact a "better" institution than Southside Virginia Community College in rural southern Virginia (mean earnings are $27,200). Thus, for colleges whose students tend to come from wealthier zip codes, we adjust the threshold upward; for those whose students come from less wealthy zip codes, we adjust the threshold down.[5]

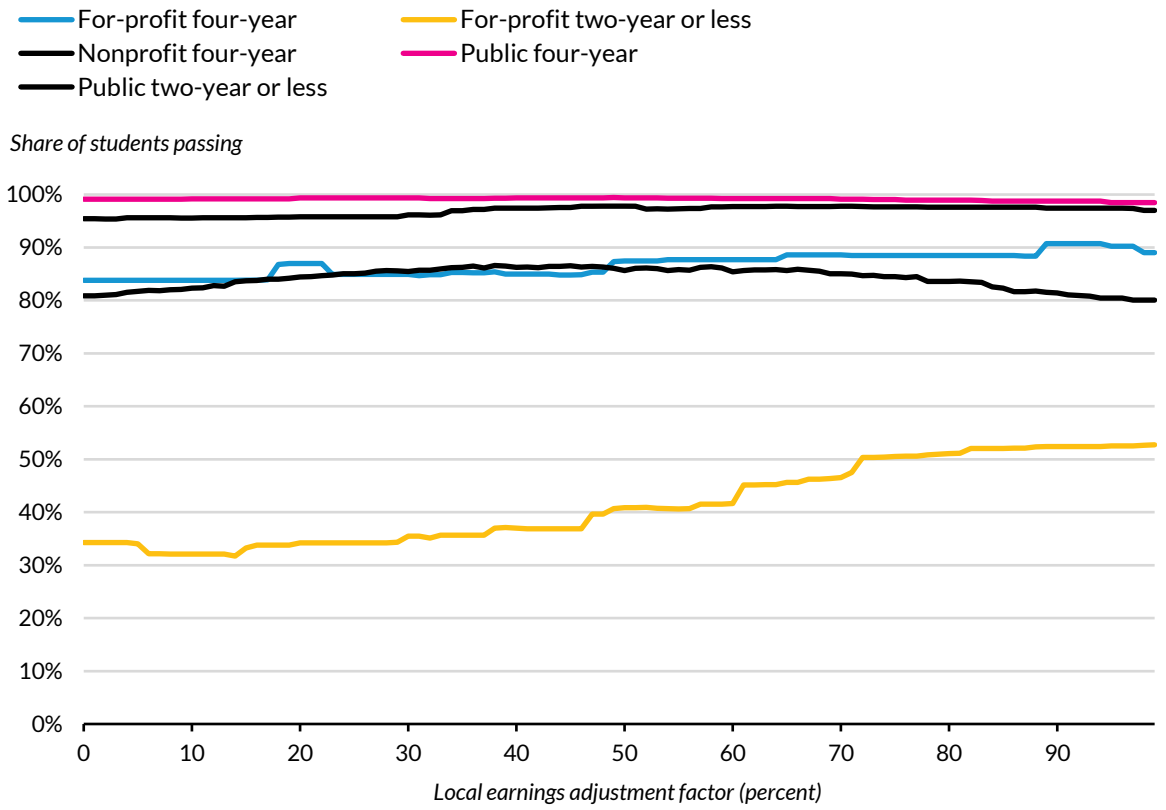
In figure 5, we explore different geographic adjustments to the income threshold. The threshold in this example is $15,000 for all colleges. Without any adjustment (i.e., with an adjustment factor of 0 percent), all institutions face the $15,000 threshold. When the adjustment factor is 100 percent, the threshold doubles to $30,000 for colleges whose students come from zip codes where earnings are twice the national average. For colleges whose students are from zip codes where earnings are half the national average, the threshold is multiplied by 0.5, falling to $7,500. Institutions in low-income areas face proportionately lower earnings thresholds. (See appendix F for a similar analysis for historically Black colleges and universities, which benefit from the geographic adjustment.)

Using only one base threshold for all sectors ($15,000), figure 5 shows that using larger geographic adjustments has no real impact on the share of students at public and private nonprofit four-year institutions attending passing institutions. But for-profit two- and less-than-two-year institutions are more likely to pass the threshold when the adjustment factor is increased (the lines on the graph slope upward as the adjustment factor increases) because their students generally come from lower-than-average-income zip codes. A geographic adjustment also has a measurable impact on historically Black colleges and universities (appendix figure F.1).

FIGURE 5

**Share of Students Attending Institutions Where at Least 75 Percent of Students Earn above a Local Threshold**

*Ten years after enrollment*




URBAN INSTITUTE

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

This is only one example of a potential adjustment for geographic differences. Policymakers may want to take a different approach, such as adjusting the earnings threshold down for low-earning zip codes but not adjusting it up for high-earning zip codes or do some other nonlinear adjustment.

### SHARE OF STUDENTS EARNING A POSITIVE RETURN ON INVESTMENT

Rather than looking merely at earnings, policymakers may wish to consider whether a student's earnings are sufficient to recoup their investment. We consider a metric similar to that proposed in the Itzkowitz study referenced above.[6] But rather than comparing the mean earnings increment with the required investment (measured as average net tuition and fees and not including forgone wages), we ask whether at least three-quarters of students earn more than they would have without college. We illustrate the estimated return on investment for various precollege earnings ("counterfactual

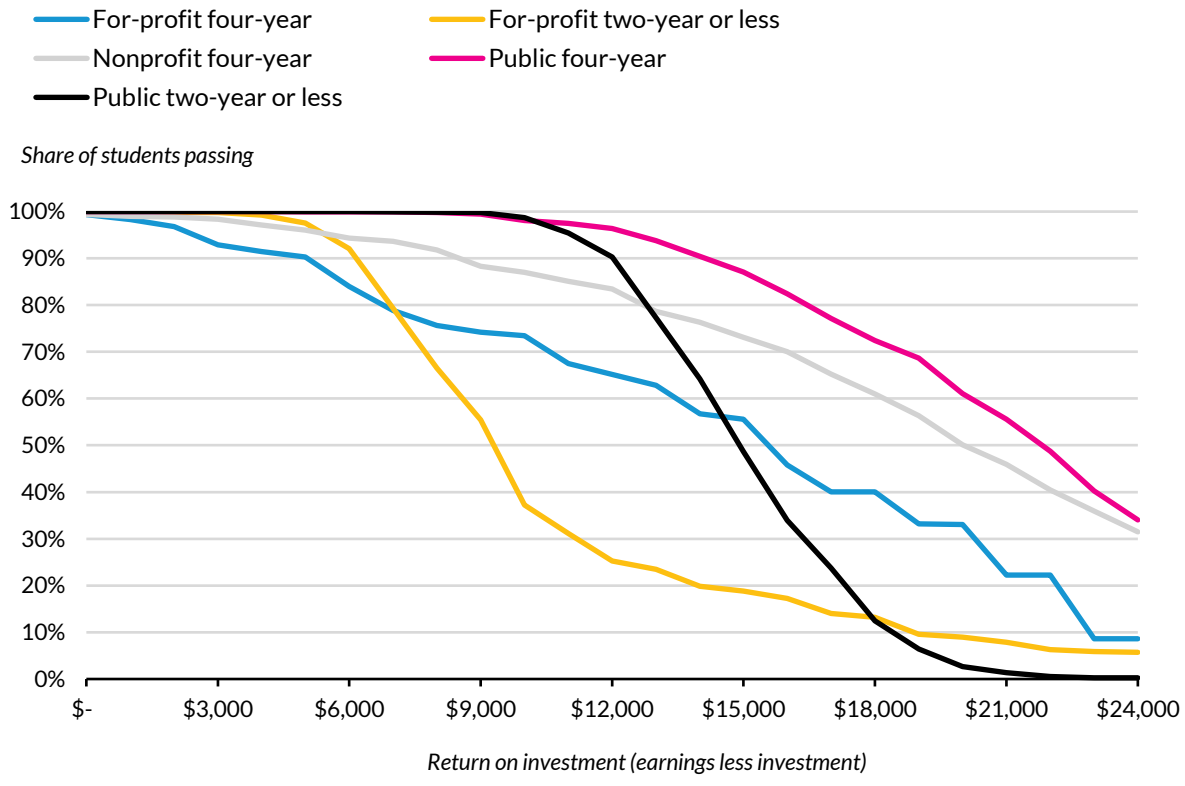
earnings") (figure 6). Because of data limitations, our metric includes all students rather than only graduates.

As the counterfactual earnings level—the earnings level assumed for those who do not go to college—increases, the share of students at institutions passing the return-on-investment test declines. As is the case with many of the earlier figures illustrating various thresholds, most institutions in all sectors pass at the lowest possible thresholds. The pass rate declines quickly for for-profit institutions and most slowly for public and private nonprofit four-year institutions. Wide differences between sectors tend to narrow at high thresholds, which few institutions in any sector pass.

Counterfactual earnings must be set very low—below the federal minimum wage for a full-time job—for most public and nonprofit institutions and any reasonable share of for-profit institutions to pass. A threshold this low would not be a credible standard for accountability.

FIGURE 6

**Return on Investment**




URBAN INSTITUTE

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

Assuming that without going to college, students would earn more than $20,000 a year leads to very few institutions outside the public and private nonprofit four-year sectors—and around half of those in these two sectors—having as many as three-quarters of their students earning enough to recoup their financial investment (figure 6).

Despite the appeal of using return on investment rather than earnings as a metric, it is impossible to recommend this approach without ensuring that both the available data and the methodology are reliable. The first step would be to include only students who complete their programs. Further investigation should probe the appearance of such dramatically low earnings outcomes.

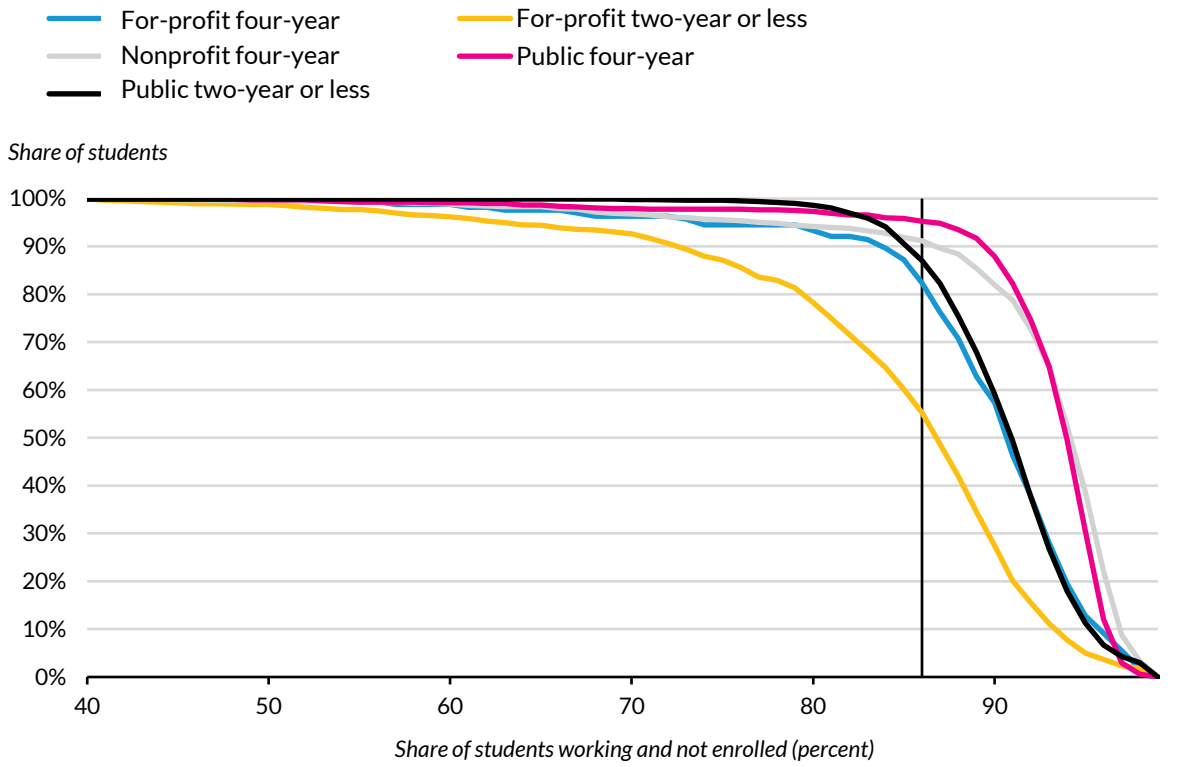
A reasonable substitute is a metric that identifies the programs and institutions that produce the largest shares of graduates with earnings below, say, the 25th percentile for similar programs.

### EMPLOYMENT RATE

A final metric considers the share of students who are employed. College Scorecard data provide the data required to consider only completers for this metric. The College Scorecard reports the number of students "working and not enrolled" and "not working and not enrolled," but it does not report a true unemployment rate, which would be the share of students who are not employed and are actively seeking work (figure 7).

FIGURE 7

**Share of Students Working and Not Enrolled Three Years after Completion**




URBAN INSTITUTE

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

# Combining Metrics

The trade-offs involved in all these choices are challenging. The specific choice of metrics will also affect how the metrics relating to different outcomes—loan repayment, completion, and labor market success—interact.

Below, we use a set of indicators to examine the potential impact of the type of multimetric system we recommend. We chose indicators based on the trade-offs, data availability, and ease of measurement.

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. Which students and institutions would be most affected by such a structure?

## Default or Repayment?

Below, we show the impact of including both a measure of default and a measure of loan repayment in the accountability system. A central question in determining whether to include two metrics related to repaying debt, but only one for labor market outcomes and one for completion, is whether this emphasis is appropriate. To pass three out of four metrics, all institutions must make a satisfactory showing on one of the debt metrics, even though it is acceptable to fail entirely in one of the other two categories. This may be problematic unless debt repayment is viewed as the best measure of quality.

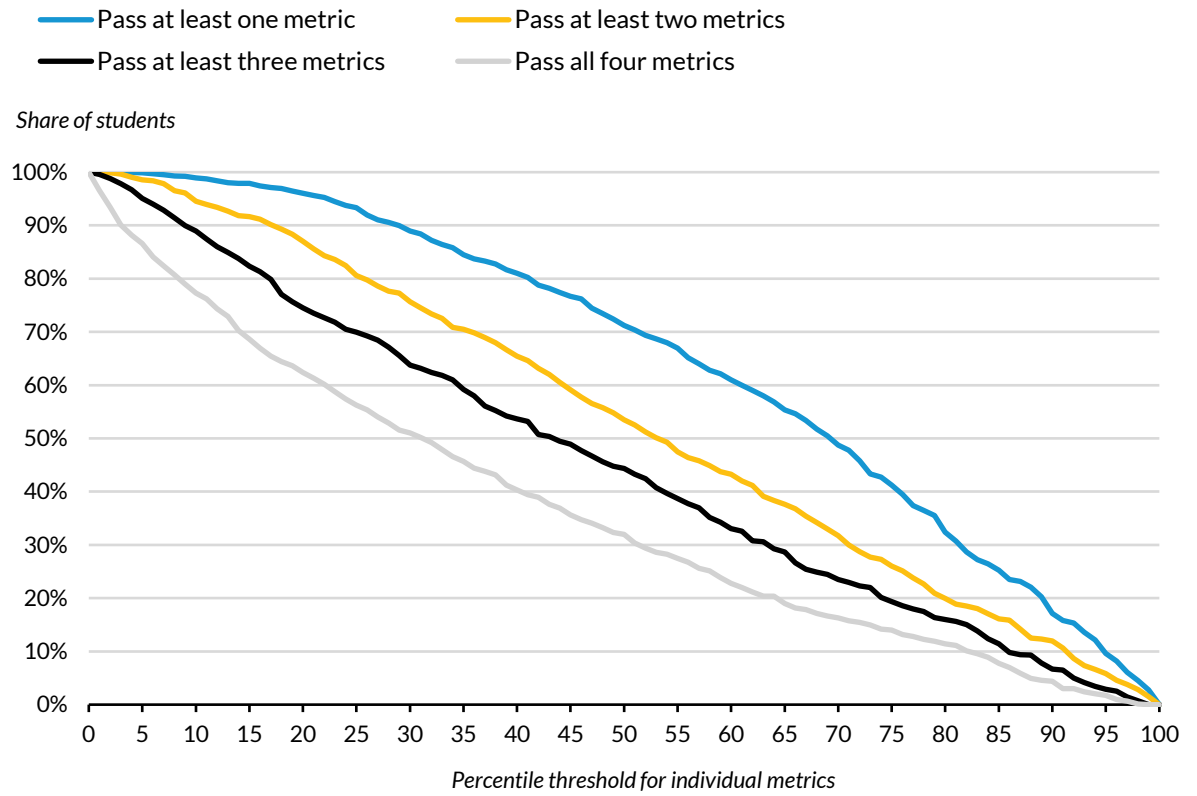
One option would be to include two metrics for completion (e.g., one that includes transfers to other institutions and one that does not) and two for labor market outcomes (e.g., earnings and unemployment), in addition to the default and repayment metrics. The difficulty with this approach is that more metrics are likely to lead to more data problems and more confusion. But it is reasonable to argue that absent clearly preferable metrics, a larger set of imperfect metrics will more accurately reflect quality.

For expositional purposes, we consider a four-metric system in which at least three metrics are needed to pass: a CDR weighted by the share of students borrowing, the share of loan dollars repaid, the completion rate after 150 percent of normal time, and the 25th percentile of earnings for all borrowers (including noncompleters). We consider different thresholds for individual metrics and ask what share of students attend institutions that pass at least one, two, three, or four metrics. (We also report the share of institutions passing.) As in previous sections, we set thresholds by percentiles so the metrics can be easily compared. For example, if the thresholds are set so that roughly 5 percent of students attend institutions that fail for each individual metric, what happens if institutions must pass multiple metrics? It turns out that if only three out of four metrics are required to pass, roughly 5 percent of students attend four-year institutions that fail (figure 8). For four-year institutions, requiring passage of all four metrics increases the failure rate to 13 percent, while requiring only one drops the failure rate to less than 1 percent.

FIGURE 8

**Share of Students Attending Institutions That Pass a Multimetric System**

*Four-year institutions*




**Source:** Authors' calculations using College Scorecard data and the Integrated Postsecondary Education Data System.
**Note:** Individual metrics are the cohort default rate adjusted for share of students borrowing, the share of loan dollars outstanding, the 150 percent completion rate, and the share of completers earning above the 25th percentile 10 years after enrollment.

When institutions fail, which metrics are they failing? Among public four-year institutions, more than half of failures to pass three out of four metrics occur because institutions fail both of the loan-related metrics or the completion and earnings metrics (table 7). But students at for-profit four-year institutions are most likely to be affected by the combination of loan repayment and completion, though this outcome describes a small number of (large) institutions. Many small institutions, enrolling very few students, tend to fail the default rate and earnings metrics.

These patterns are different for two-year institutions (from all sectors combined), where the CDR and earnings cause the plurality of failures (appendix table G.1), and for less-than-two-year institutions, where failures are predominantly caused by loan repayment and completion (appendix table G.2). These differences occur even though the thresholds for each metric are set separately for different types of

institutions with, for example, 5 percent of students in both two-year and four-year institutions attending failing institutions for each metric. In table 7, the share of institutions failing different combinations of metrics differs from the share of student failing when the patterns differ for large and small institutions.

TABLE 7

**Causes of Failure in a Multimetric System**

*Four-year institutions, by sector*

| | Share of Institutions Failing | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Public Four-Year | | Private Nonprofit Four-Year | | Private For-Profit Four-Year | | All Four-Year | |
| | Institutions | Students | Institutions | Students | Institutions | Students | Institutions | Students |
| Cohort default rate, loan repayment | 3% | 1% | 1% | 1% | 3% | 1% | 2% | 1% |
| Loan repayment, completion | 0% | 0% | 0% | 0% | 2% | 15% | 0% | 1% |
| Completion, earnings | 4% | 1% | 1% | 0% | 1% | 0% | 2% | 1% |
| Cohort default rate, loan repayment, completion | 0% | 0% | 0% | 0% | 3% | 8% | 1% | 1% |
| Cohort default rate, loan repayment, earnings | 1% | 0% | 2% | 0% | 6% | 3% | 2% | 0% |
| Cohort default rate, earnings | 0% | 0% | 1% | 0% | 14% | 6% | 2% | 0% |
| Cohort default rate, completion, earnings | 0% | 0% | 0% | 1% | 2% | 0% | 0% | 0% |
| Loan repayment, earnings | 0% | 0% | 1% | 0% | 6% | 1% | 1% | 0% |
| Cohort default rate, completion | 0% | 0% | 0% | 0% | 4% | 2% | 0% | 0% |
| Loan repayment, completion, earnings | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% |
| Cohort default rate, loan repayment, completion, earnings | 0% | 0% | 1% | 0% | 2% | 0% | 1% | 0% |
| Total | 10% | 3% | 7% | 3% | 43% | 37% | 10% | 5% |
| Total number failing | 62 | 250,228 | 73 | 94,346 | 46 | 223,176 | 181 | 567,750 |

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

Of course, policymakers may wish to set different thresholds for each metric; there is no reason for the relative threshold to be the same across all metrics. Moreover, policymakers may wish to consider additional metrics. More metrics means a stronger accountability system and one that is less conducive to gaming but introduces more oversight, bureaucracy, and room for error.

# Conclusion

We have argued for some basic choices in developing an accountability system:

- **One metric or multiple metrics?** The system should rely on multiple metrics to avoid placing too much weight on one imperfect measure. Programs and institutions have different strengths and weakness, and using multiple metrics accommodates this diversity.
- **Meet all metrics or a subset?** The system should incorporate flexibility, requiring programs or institutions to meet three out of four (or four out of five or two out of three) well-designed metrics. Such a system is appropriate because the metrics are imperfect proxies for quality and because the data are frequently flawed. We use the example of three out of four.
- **Set separate benchmarks for different types of institutions?** 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 specific standards that will apply equally well to all types of programs and institutions. Earnings following bachelor's degree programs must be higher than those following shorter programs to generate a reasonable rate of return. And the shorter time requirement for certificates as opposed to associate degrees makes it more likely that students facing challenges will complete the less demanding programs. Setting one earnings threshold or one completion rate threshold for all programs and institutions will, therefore, create a biased system.

The outcomes we documented in this report 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. To isolate 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 the vast majority of students. Setting standards will provide incentives for improved performance, in addition to eliminating 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 with the resources and strategies required to support student success.

# Appendix A. Overview of the Existing Regulatory System

## Default Rates

For schools with 30 or more borrowers entering repayment in a fiscal year, the school's CDR is the share of a school's borrowers who enter repayment during that fiscal year and default within three years. For schools with 29 or fewer borrowers entering repayment during a fiscal year, the CDR is based on borrowers entering repayment over a three-year period.

The default rate does not include PLUS loans made to parents of dependent students or to graduate students (or Perkins loans).

If a school's three-year CDR is 30 percent or higher for three consecutive years or higher than 40 percent for one year, the school loses eligibility for Pell grants and direct loans for three years (FSA 2020). Only four institutions had default rates of 30 percent or higher in fiscal years 2016, 2017, and 2018. Two of these four, in addition to four other institutions, had default rates greater than 40 percent in fiscal year 2018.[7]

## Financial Responsibility Composite Scores

The Department of Education uses a composite of three measures derived from institutions' audited financial statements to monitor institutional financial health: a primary reserve ratio (which measures a school's viability and liquidity), an equity ratio (which measures its capital resources and ability to borrow), and a net income ratio (which measures its profitability). The scores reflect the extent to which a school has the financial resources to (1) replace existing technology with newer technology; (2) replace physical capital that wears out; (3) recruit, retain, and retrain faculty and staff (human capital); and (4) develop new programs (FSA 2021).

The scale ranges from -1.0 to 3.0, with scores of 1.5 or higher interpreted as indicating sound financial circumstances. Schools scoring between 1.0 and 1.5 are subject to additional oversight.

Schools scoring below 1.0 are considered not financially responsible but may continue to participate in Title IV programs under provisional certification. Schools with these low scores are generally subject to cash monitoring requirements and must post a letter of credit equal to a minimum of 10 percent of the Title IV aid received in the most recent fiscal year. Larger letters of credit can relieve the institution of other penalties.

In addition to the composite score, the Department of Education looks at the cash reserves needed to make required refunds and confirms that schools are meeting their financial obligations, including making debt payments. Both for-profit and private nonprofit institutions are subject to these rules.

## 90/10 Rule

In 1992, Congress enacted a rule limiting the share of revenues for-profit institutions participating in Title IV programs could receive from federal student aid to 85 percent. The required ratio was changed to 90 percent / 10 percent in 1998. This standard applies only to for-profit institutions (and those converting to nonprofit status for one year beyond that change). The 2021 American Rescue Plan Act modified the 90/10 rule to count aid to military personnel and veterans along with other federal financial aid in the 90 percent beginning in 2023.

The rule emerged from for-profit institutions springing up specifically to serve veterans after World War II.

Misgivings about the rule include arguments that it penalizes schools based on their students' financial need rather than on quality, that it leads to tuition increases to ensure federal aid cannot cover the full charges, and that it is too weak.

Lee and Looney (2019) argue that almost all public and private nonprofit institutions would pass this test, so it would not be productive to apply it to them. But Kantrowitz (2013) argues that state appropriations should be included in the 90 percent because these funds substitute for the tuition revenues that support for-profit institutions.

# Gainful Employment

The Gainful Employment Rule, adopted by the Department of Education in 2014 and rescinded in 2019 before it was implemented, set standards for all programs at for-profit institutions and nondegree programs at nonprofit institutions.

The rule required any program where typical graduates' debts exceed both 8 percent of their total income and 20 percent of their discretionary income to improve or lose access to federal financial aid.

The numerator is the calculated annual loan payment amount, an estimate of the annual loan repayment amount based on the median educational debt of the cohort members. The denominator is the higher of the cohort's mean or median earnings obtained from the Social Security Administration. Ratios less than or equal to 8 percent are passing rates. Those between 8 percent and 12 percent are in a warning zone, and ratios above 12 percent are failing rates. For the discretionary income rate, the denominator is the higher of the mean or median earnings minus 150 percent of the federal poverty level. Ratios less than or equal to 20 percent are passing rates, rates greater than 20 percent but less than or equal to 30 percent are warning zone rates, and rates greater than 30 percent are failing rates.[8] An earlier version of the regulations included a threshold for the share of students (including completers and noncompleters) who are successfully repaying their loans, but a court objected to this provision, arguing that there was inadequate evidence for setting the threshold.

Concerns about the adequacy of the gainful employment regulations include their exclusive focus on program graduates and the applicability only to a subset of programs and institutions.

# Appendix B. Choosing a Time Period

## Dollar-Based Repayment Rates

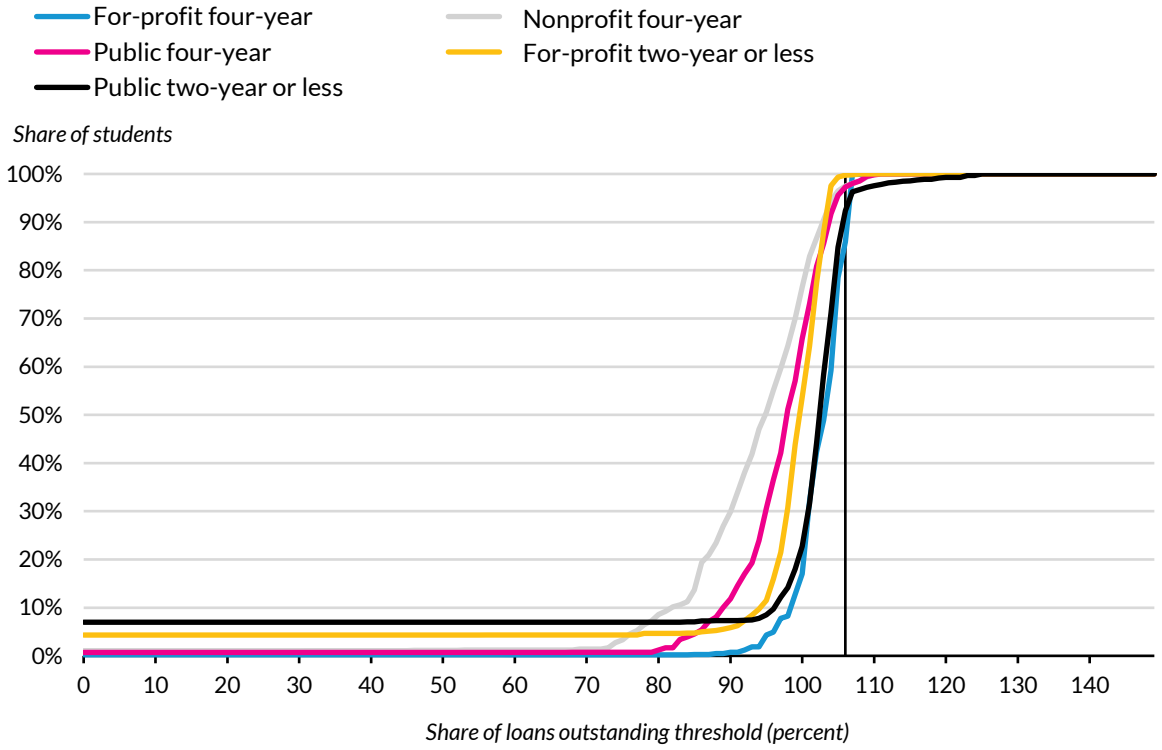
In the main text, we report on the share of debt outstanding after five years (figure 2; tables 4 and 5). Data are also available for 1, 4, 10, and 20 years after entering repayment. These measures are available separately for undergraduate, graduate, and Parent PLUS loans, enabling separate accountability rules for undergraduate and graduate programs.

As with many of the other metrics, one of the decision points here is when to measure loan repayment. Measuring soon after leaving allows the system to work more quickly to impose sanctions, but a longer time frame allows time for earnings to stabilize. In practice, 1 year is too soon to differentiate between institutions (figure B.1), while 20 years is impractical for accountability. Five years appears to be sufficient time for noticeable patterns to emerge. Figure B.2 shows 5-year dollar-based repayment rates for graduate students.

FIGURE B.1

**Share of Undergraduate Students Attending Institutions Passing the Dollar-Based Repayment Metric after One Year**

*Share of dollars outstanding*



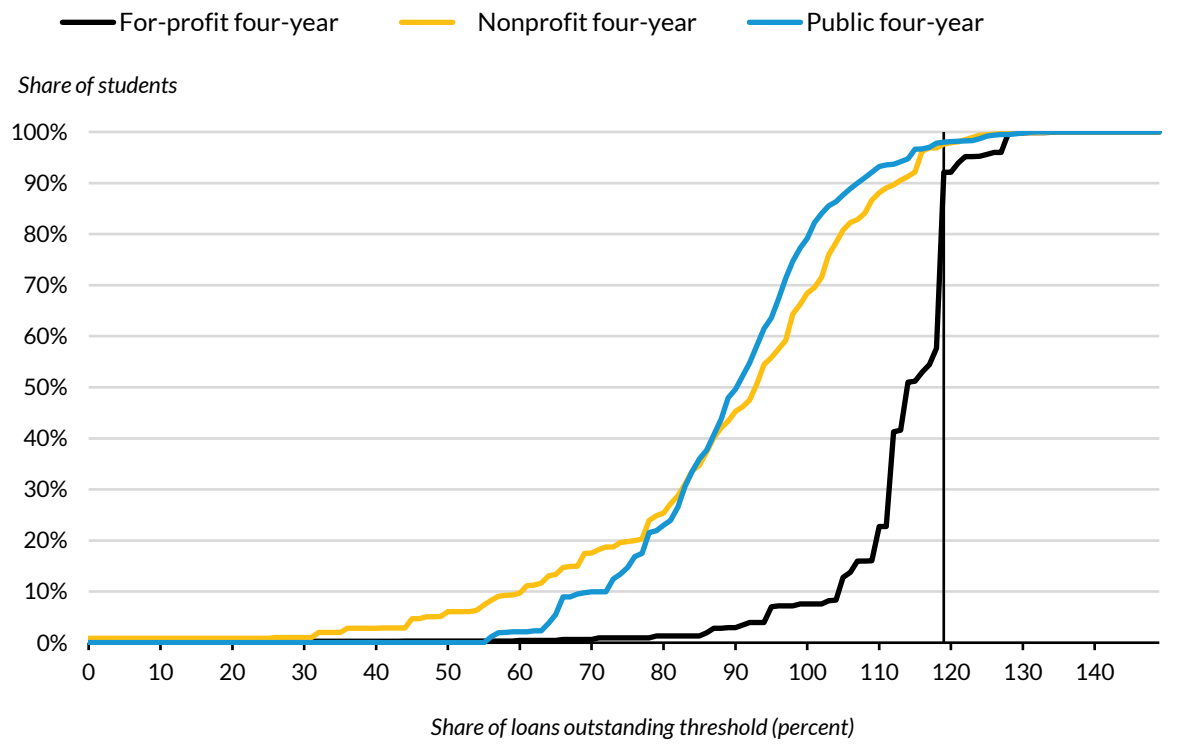

URBAN INSTITUTE

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

FIGURE B.2

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

*Share of dollars outstanding*




URBAN INSTITUTE

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

# Appendix C. Alternative Metrics

In this appendix, we describe alternative metrics to those described in the body of the report.

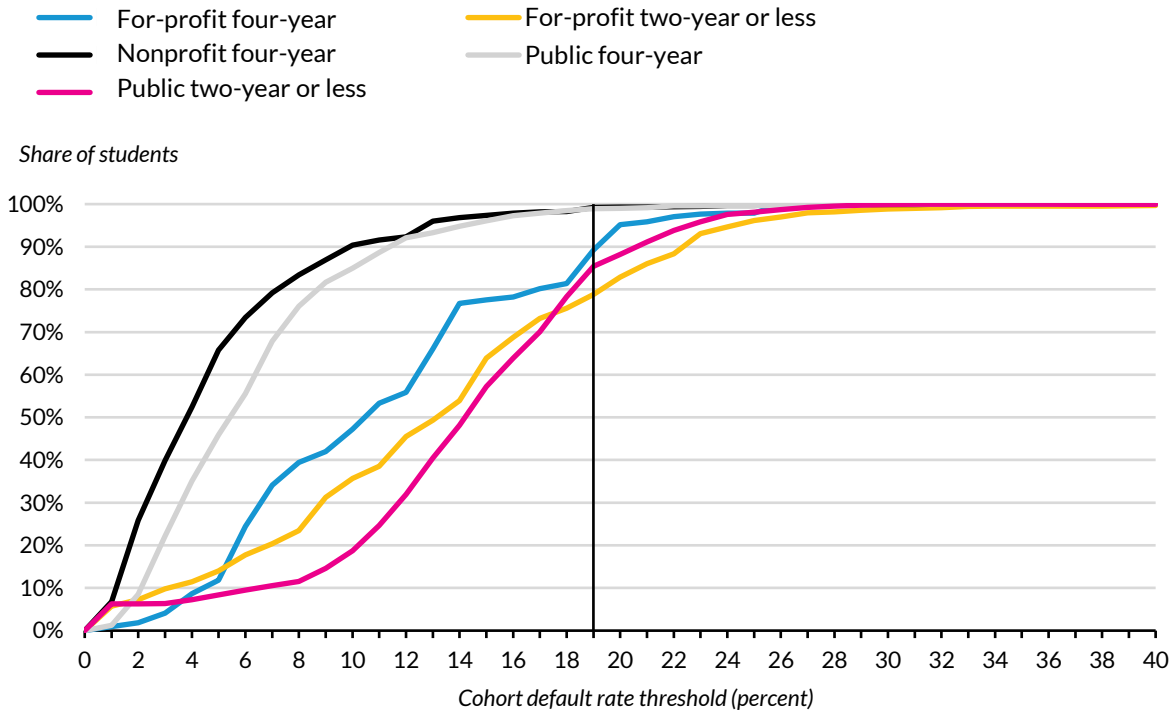
## Cohort Default Rate

Figure 1 in the main text illustrates a metric based on CDRs weighted by the share of students borrowing. Figure C.1 relies on CDRs that are not affected by whether most students or only a small share take out student loans. The most salient difference is that with the unweighted metric, public two-year colleges have one of the lowest passing rates, whereas with the weighted metric, these institutions perform well because a relatively small share of their students borrows.

FIGURE C.1

**Share of Students Attending Institutions Passing the Unweighted Cohort Default Rate Metric**

*Cohort default rate 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 vertical line represents the 95th percentile cohort default rate (institutions weighted by full-time equivalent students).

TABLE C.1

**Unweighted Cohort Default Rate above 19 Percent**

| | Students attending failing institutions | All students | Share of students in sector attending failing institutions | Distribution of students attending failing institutions, by sector |
|---|---|---|---|---|
| For-profit two-year | 55,637 | 266,163 | 21% | 7% |
| For-profit four-year | 85,720 | 793,165 | 11% | 11% |
| For-profit less-than-two-year | 39,203 | 204,355 | 19% | 5% |
| Nonprofit two-year | 4,979 | 44,852 | 11% | 1% |
| Nonprofit four-year | 29,093 | 3,613,508 | 1% | 4% |
| Nonprofit less-than-two-year | 863 | 7,030 | 12% | 0% |
| Public two-year | 508,036 | 3,580,631 | 14% | 63% |
| Public four-year | 81,959 | 7,615,539 | 1% | 10% |
| Public less-than-two-year | 2,050 | 42,749 | 5% | 0% |
| **Total** | **807,540** | **16,167,992** | **5%** | **100%** |

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

# Borrower-Based Repayment Rates

The repayment rate metric detailed in the body of the report is based on the share of dollars remaining. In contrast, a borrower-based repayment rate focuses on the share of borrowers who have reduced their outstanding balances.

In the for-profit and community college sectors, fewer than half of borrowers have paid down even a dollar on their principal after five years. These numbers rise when weighted by the share of borrowers because they are based on the share of students, rather than the share of borrowers who have debt they have not begun to retire. As in the case of default rates, this adjustment is particularly dramatic among community colleges, where proportionally fewer students borrow.

TABLE C.2

**Average Five-Year Repayment Rates**

*Shares of students who have reduced their loan principal by at least one dollar*

| | Five-year repayment rate | Five-year repayment rate, adjusted by share of students borrowing | Five-year repayment rate, weighted by number of borrowers | Weighted five-year repayment rate, adjusted by share of students borrowing |
|---|---|---|---|---|
| For-profit two-year | 41% | 61% | 37% | 57% |
| For-profit four-year | 43% | 64% | 37% | 60% |
| For-profit less-than-two-year | 42% | 65% | 38% | 59% |
| Nonprofit two-year | 56% | 71% | 39% | 48% |
| Nonprofit four-year | 67% | 79% | 68% | 80% |
| Nonprofit less-than-two-year | 51% | 70% | 40% | 65% |
| Public two-year | 45% | 89% | 43% | 86% |
| Public four-year | 61% | 83% | 65% | 83% |
| Public less-than-two-year | 51% | 69% | 49% | 66% |
| **Total** | 54% | 77% | 60% | 80% |

**Source:** Authors' calculations using 2016–17 College Scorecard data.

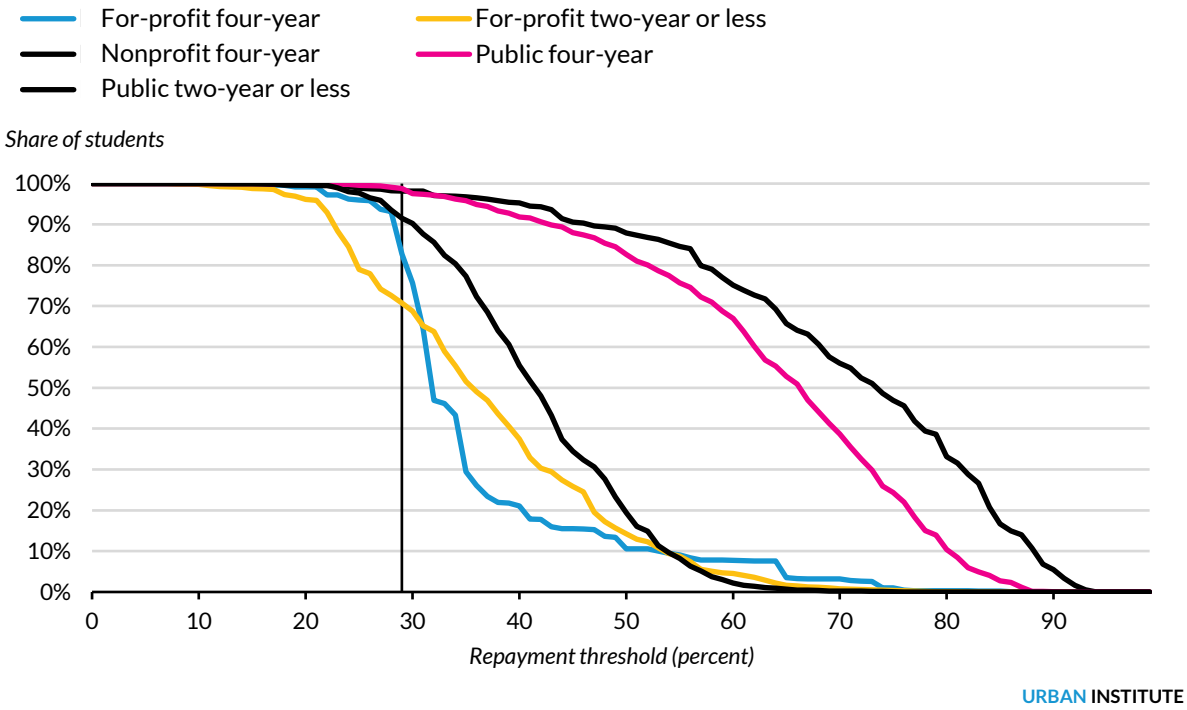
Figure C.2 shows the effects of setting thresholds at different levels for the unweighted borrower-based repayment rate. The 5th percentile of repayment rates is 30 percent; table C.3 shows the shares of students in each sector attending institutions that would fail such a threshold. (Note that the 30 percent falls on a particularly steep part of the curve for four-year for-profits. In practice, we would advise against setting a threshold here, instead recommending a value below 27 percent or above 40 percent to avoid having the futures of too many institutions and students rest on small changes in outcomes.)

Figure C.3 shows the outcome when the repayment rate is weighted by the share of students borrowing. Relative to the unweighted repayment rate, for-profit institutions (where a large share of students borrow) tend to do worse with the weighted repayment rate and community colleges (where a small share of student borrow) do better, suggesting that borrowers at for-profit institutions are making less progress on their loans than even their poor default rates suggest. This could be because for-profit institutions are better at steering their students into forbearance or deferment, where their balances tend to increase because of unpaid interest and where borrowers do not make progress toward loan forgiveness. Some institutions may also encourage income-driven repayment plans, which mitigate the possibility of default but do not help students make progress on their loans if the required payment does not cover interest.

FIGURE C.2

**Share of Students Attending Institutions Passing the Repayment Rate Metric after Five Years, Based on Share of Students Reducing Their Loan Balance by at Least One Dollar**

*Repayment rate thresholds from 0 to 100 percent*




URBAN INSTITUTE

**Source:** Authors' calculations using 2016–17 College Scorecard data and the Integrated Postsecondary Education Data System.

TABLE C.3

**Share of Students Attending Institutions Where Less Than 30 Percent of Borrowers Have Reduced Their Loan Principal by at Least One Dollar after Five Years**

| | Students attending failing institutions | All students | Share of students in sector attending failing institutions | Distribution of students attending failing institutions by sector |
|---|---|---|---|---|
| For-profit two-year | 75,862 | 255,367 | 30% | 9% |
| For-profit four-year | 174,005 | 779,797 | 22% | 20% |
| For-profit less-than-two-year | 52,806 | 165,616 | 32% | 6% |
| Nonprofit two-year | 19,725 | 40,009 | 49% | 2% |
| Nonprofit four-year | 63,405 | 3,516,053 | 2% | 7% |
| Nonprofit less-than-two-year | 2,373 | 4,913 | 48% | 0% |
| Public two-year | 286,813 | 3,124,503 | 9% | 34% |
| Public four-year | 177,058 | 7,526,188 | 2% | 21% |
| Public less-than-two-year | 101 | 13,474 | 1% | 0% |
| **Total** | **852,148** | **15,425,920** | **6%** | **100%** |

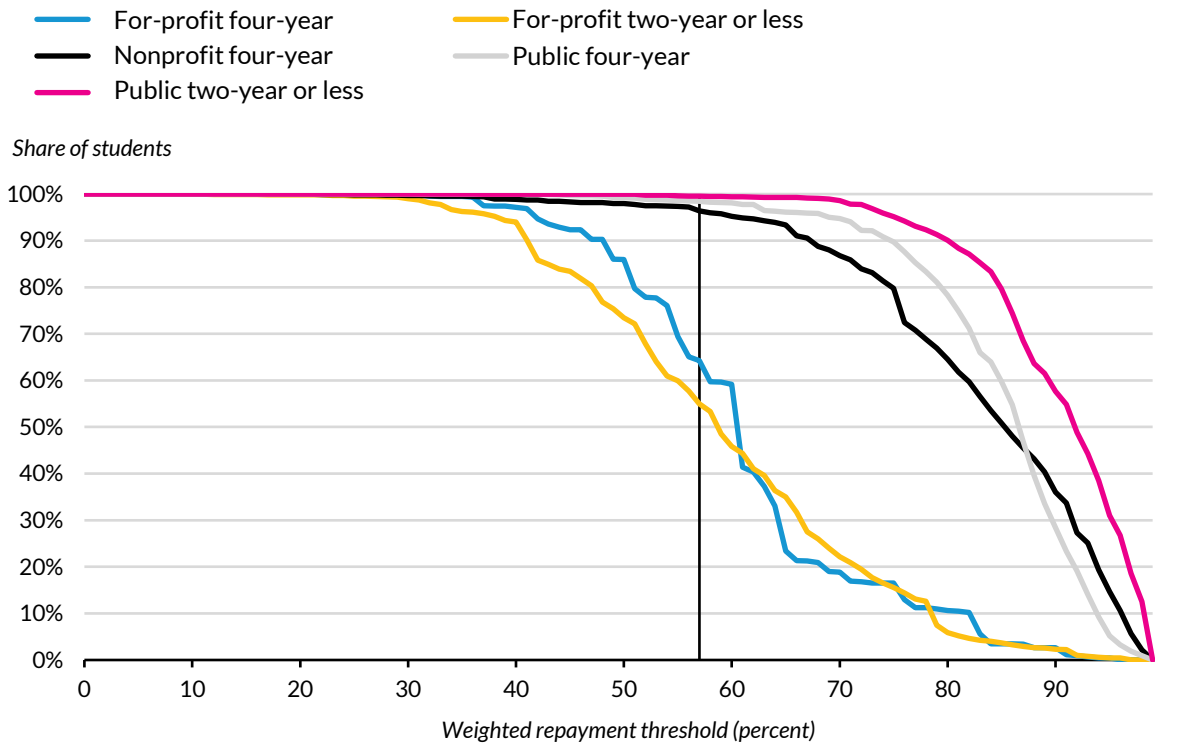
**Source:** Authors' calculations using 2016–17 College Scorecard data and the Integrated Postsecondary Education Data System.

Table C.4 shows that with the weighted repayment rate, community college students compose only 2 percent of students at failing institutions, with over half at for-profit institutions. In fact, nearly half the institutions in the for-profit sector fail the weighted repayment rate metric when set at the 5th percentile, reflecting the fact that large shares of their students borrow and fail to succeed at repaying their debts.

FIGURE C.3

**Share of Students Attending Institutions Passing the Weighted Repayment Rate Metric after Five Years, Based on Share of Students Reducing Their Loan Balance by at Least One Dollar**

*Thresholds from 0 to 100 percent*




URBAN INSTITUTE

**Source:** Authors' calculations using 2016–17 College Scorecard data and the Integrated Postsecondary Education Data System.

TABLE C.4

**Weighted Repayment Rate**

*Shares of institutions and students failing*

| | Students attending failing institutions | All students | Share of sector | Share of failures |
|---|---|---|---|---|
| For-profit two-year | 119,636 | 255,367 | 47% | 15% |
| For-profit four-year | 303,516 | 779,797 | 39% | 39% |
| For-profit less-than-two-year | 70,494 | 165,616 | 43% | 9% |
| Nonprofit two-year | 21,888 | 40,009 | 55% | 3% |
| Nonprofit four-year | 127,400 | 3,516,053 | 4% | 16% |
| Nonprofit less-than-two-year | 1,222 | 4,913 | 25% | 0% |
| Public two-year | 12,104 | 3,124,503 | 0% | 2% |
| Public four-year | 118,200 | 7,526,188 | 2% | 15% |
| Public less-than-two-year | 2,894 | 13,474 | 21% | 0% |
| **Total** | **777,354** | **15,425,920** | **5%** | **100%** |

**Source:** Authors' calculations using 2016–17 College Scorecard data and the Integrated Postsecondary Education Data System.

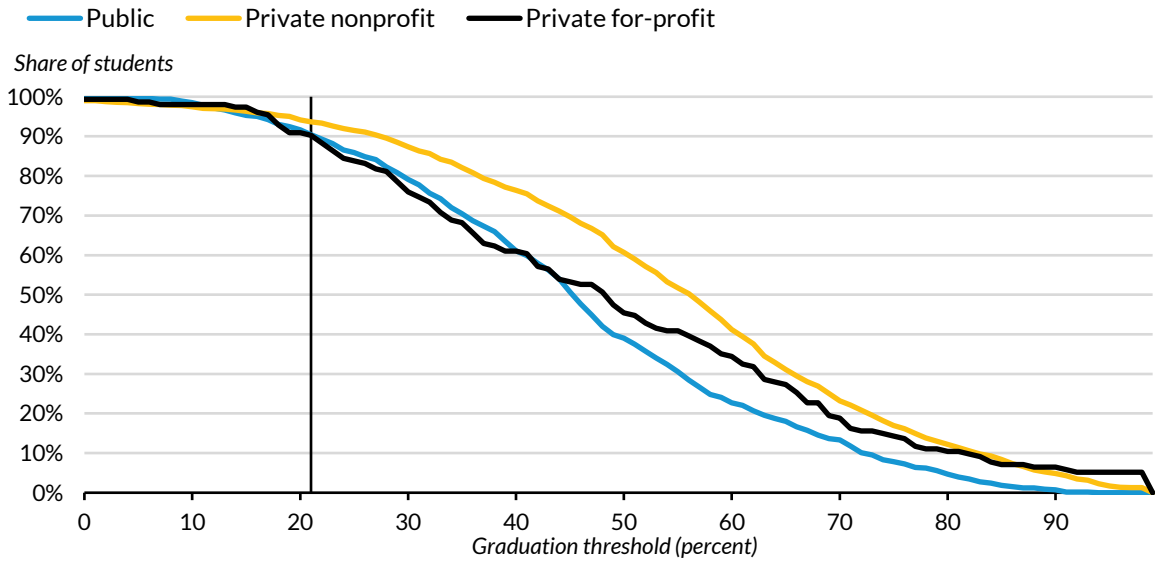
# Appendix D. Completion Rates

Figure 3 in the main text shows how different types of institutions perform on the metric based on the share of first-time full-time students completing their programs within 150 percent of the normal time. Figures D.1, D.2, and D.3 show the impact of setting different thresholds for four-year, two-year, and less-than-two-year institutions.

FIGURE D.1

**Share of Students Attending Institutions with Completion Rates above a Specified Level**

*Four-year institutions*



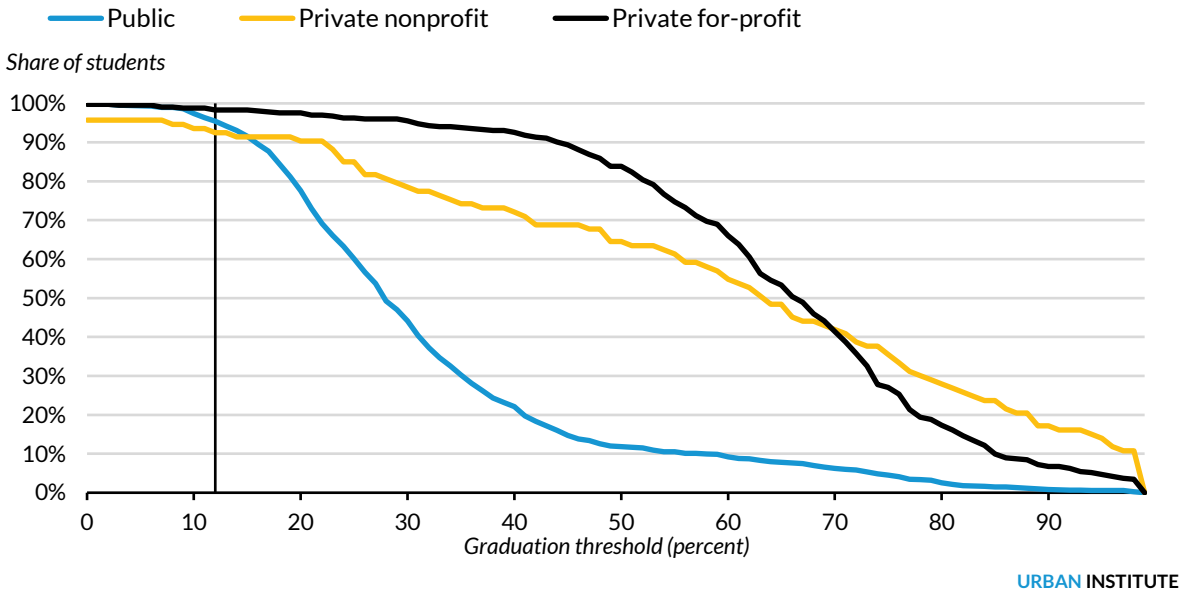

URBAN INSTITUTE

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

FIGURE D.2

**Share of Students Attending Institutions with Completion Rates above a Specified Level**

*Two-year institutions*

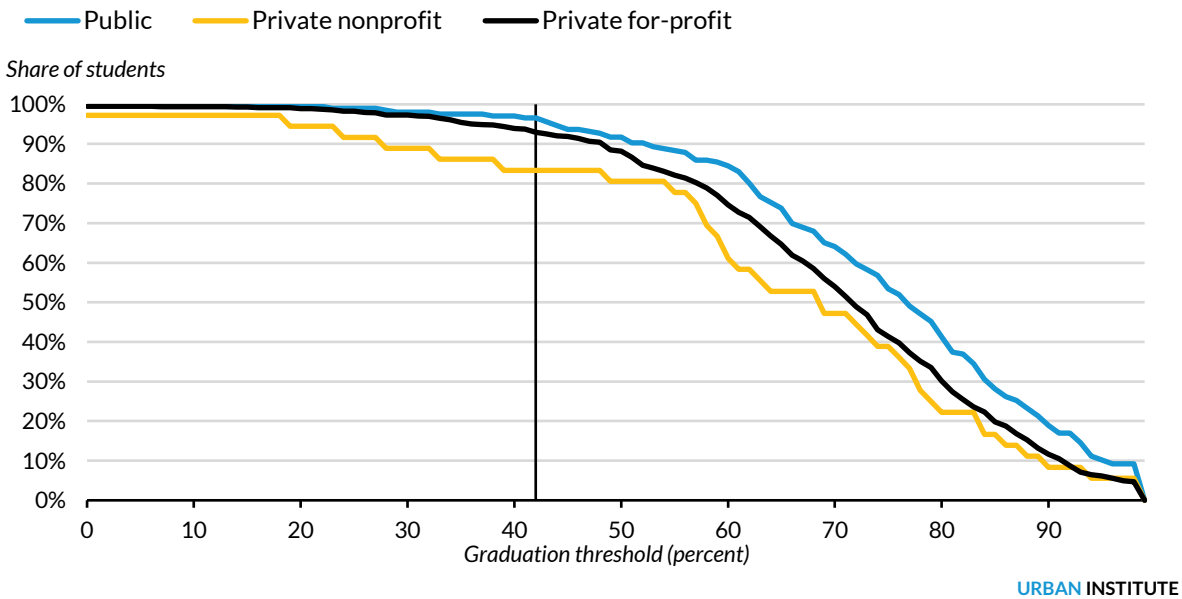



**Source:** Authors' calculations using 2019–20 College Scorecard data and the Integrated Postsecondary Education Data System.

FIGURE D.3

**Share of Students Attending Institutions with Completion Rates above a Specified Level**

*Less-than-two-year institutions*




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

## Part-Time Students

Some institutions serve large shares of part-time students and should arguably be judged on outcomes for these students, rather than only on first-time, full-time students. But with some exceptions for very low thresholds, it is unclear whether using a metric for part-time students alters the picture substantially. About 13 percent of community college students attend institutions where the completion rate for part-time students within six years is higher than the standard IPEDS completion rate. Furthermore, less-than-two-year institutions do not (currently) report this metric, limiting its usefulness for a large swath of for-profit institutions.

## Transfer Students

Concerns about students transferring out and completing degrees raise the question of using a metric that considers the share of students completing at *any* institution within six years. But in the College Scorecard data, this metric is available only for students receiving Title IV federal student aid and only if they were Title IV students at both institutions (enabling the linkage). What continues to stand out is how poorly community colleges do, despite the alleged mission of many to prepare students to transfer to four-year institutions. On average, community colleges reporting both metrics average an IPEDS 150 percent graduation rate of 26 percent and a graduation-anywhere rate of 33 percent.

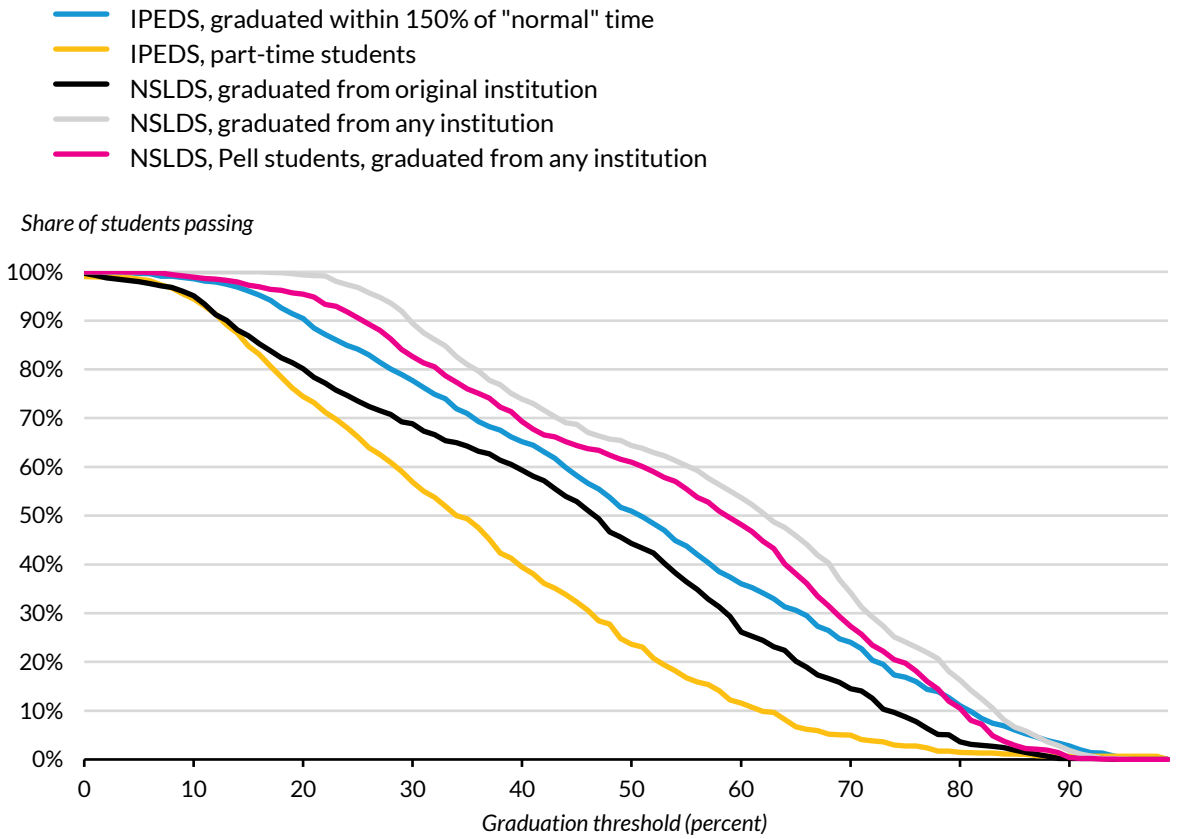
## Pell Grant Recipients

Finally, one might wish to consider low-income students or Pell grant recipients separately. Providing institutions incentives to support their most at-risk students to complete their programs is an important goal, but many institutions do not report this metric. To gain insight into the implications of the choice of a graduation rate definition, figure D.4 compares five graduation rates for the institutions that report all of them.

Comparisons among the metrics are problematic because of the large share of institutions that do not provide completion information. Overall, including graduation from any institution improves the pass rate, but these data are available only for students receiving federal student aid. The lowest pass rate is for the IPEDS part-time completion rate.

FIGURE D.4

**Alternate Graduation Rates**




URBAN INSTITUTE

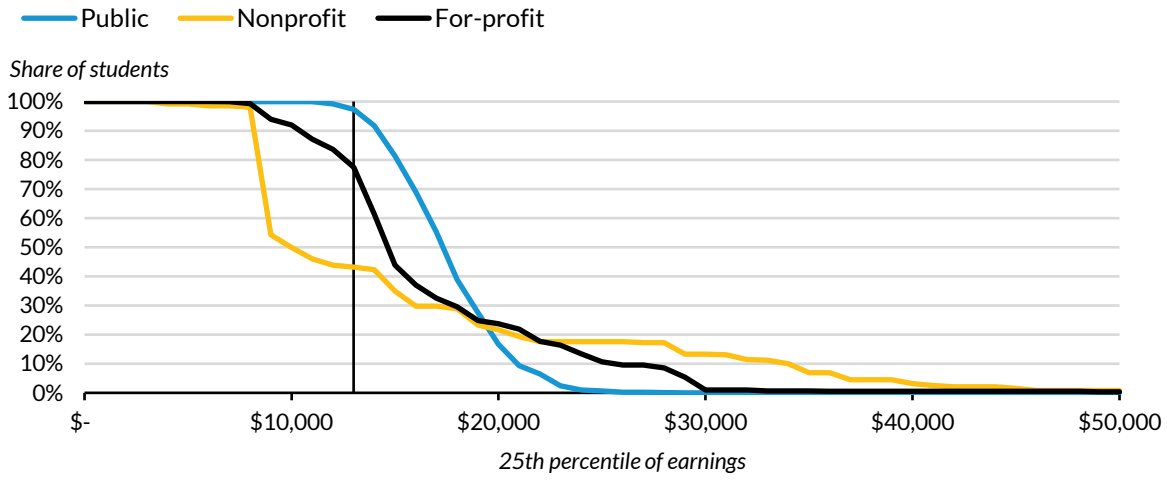
**Source:** Authors' calculations using 2019–20 College Scorecard data and the Integrated Postsecondary Education Data System.
**Note:** IPEDS = Integrated Postsecondary Education Data System; NSLDS = National Student Loan Data System.

# Appendix E. Earnings

FIGURE E.1

**Share of Students Attending Institutions Where at Least 75 Percent Earn Specified Levels 10 Years after Enrolling**

*Two-year institutions*

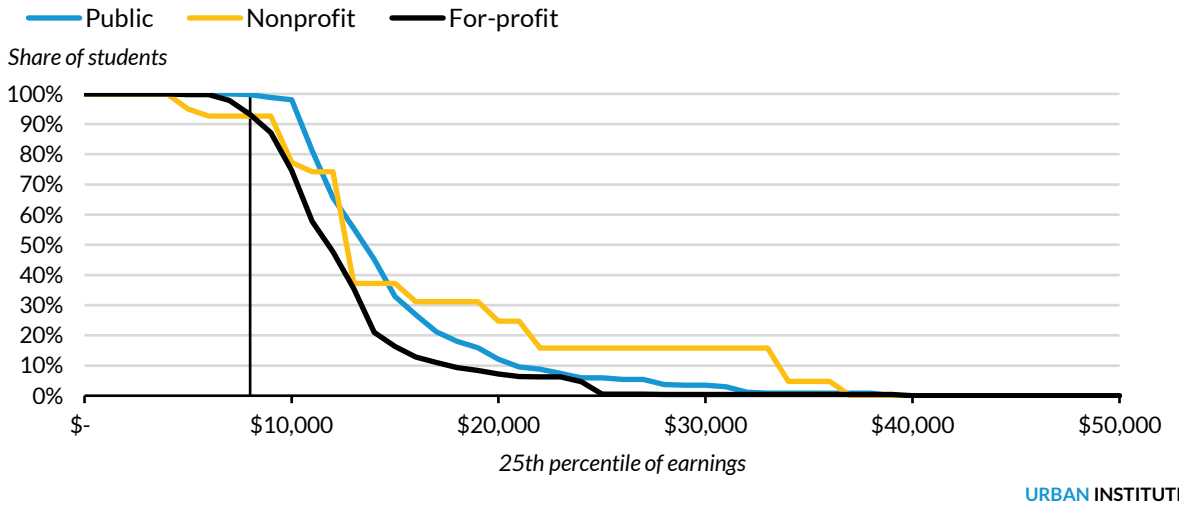



**Source:** Authors' calculations using 2014–15 College Scorecard data and the Integrated Postsecondary Education Data System.

FIGURE E.2

**Share of Students Attending Institutions Where at Least 75 Percent Earn Specified Levels 10 Years after Enrolling**

*Less-than-two-year institutions*




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

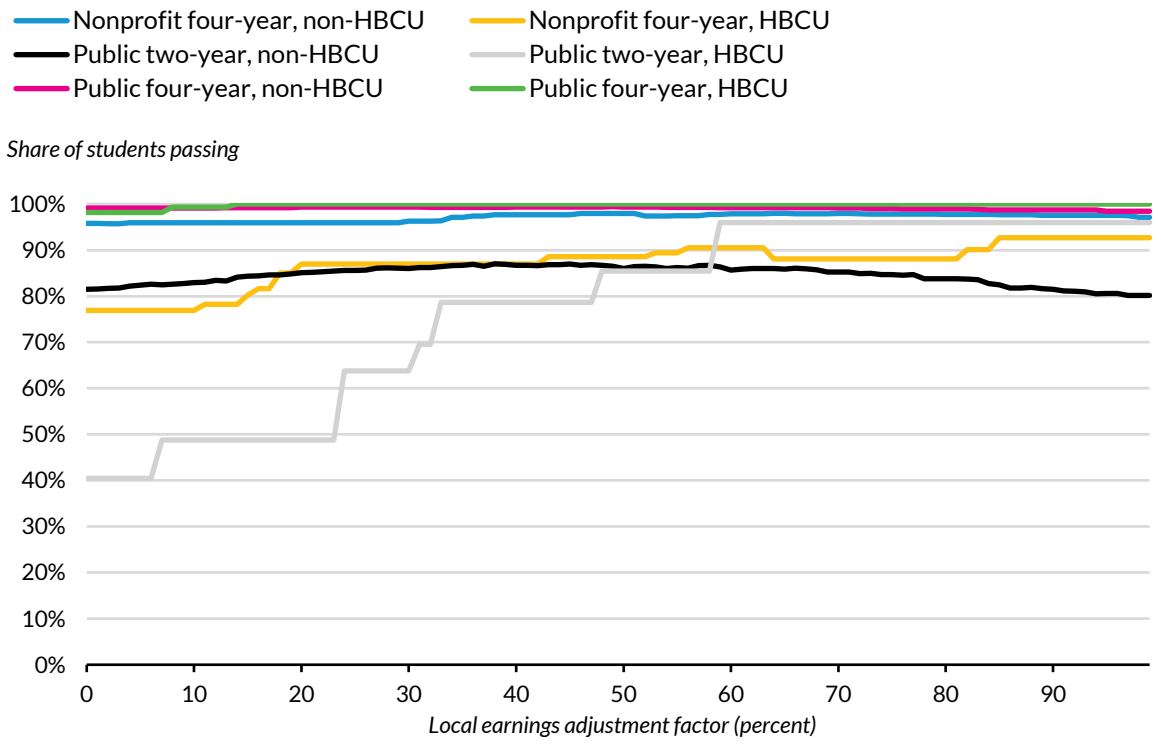
# Appendix F. Geographic Adjustment for HBCUs

The geographic adjustment illustrated in the body of the report (figure 5) shows that adjusting the earnings threshold down for schools enrolling students from low-income areas improves the performance of two-year and less-than-two-year institutions. Similarly, this adjustment has a significant impact on HBCUs. Because their students tend to come from low-income areas, their earnings thresholds adjust down, and more institutions pass than without this adjustment.

FIGURE F.1

**Share of Students Attending Institutions Where at Least 75 Percent of Students Earn above a Local Threshold**

*HBCUs, 10 years after enrollment*




URBAN INSTITUTE

**Source:** Authors' calculations using 2014–15 College Scorecard data and the Integrated Postsecondary Education Data System.
**Note:** HBCU = historically Black college and university.

# Appendix G. Multimetric System

TABLE G.1

**Causes of Failure in a Multimetric System**

*Two-year institutions*

| | Share of institutions failing | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Public | | Private Nonprofit | | Private For-Profit | | All | |
| Failed metrics | Institutions | Students | Institutions | Students | Institutions | Students | Institutions | Students |
| Cohort default rate, earnings | 0% | 0% | 2% | 43% | 8% | 7% | 2% | 1% |
| Cohort default rate, loan repayment | 0% | 0% | 7% | 3% | 6% | 5% | 2% | 0% |
| Loan repayment, completion | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% |
| Cohort default rate, completion | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% |
| Cohort default rate, loan repayment, earnings | 0% | 0% | 0% | 0% | 3% | 1% | 1% | 0% |
| Loan repayment, earnings | 0% | 0% | 4% | 0% | 2% | 1% | 1% | 0% |
| Completion, earnings | 1% | 0% | 4% | 1% | 0% | 0% | 1% | 0% |
| Cohort default rate, loan repayment, completion | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% |
| Cohort default rate, loan repayment, completion, earnings | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% |
| Total | 2% | 1% | 18% | 48% | 20% | 15% | 6% | 2% |
| Total number failing | 17 | 40,934 | 8 | 18,025 | 43 | 31,523 | 68 | 90,482 |

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

TABLE G.2

**Causes of Failure in a Multimetric System**

*Less-than-two-year institutions*

| | Share of institutions failing | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Public | | Private Nonprofit | | Private For-Profit | | All | |
| **Failed metrics** | **Institutions** | **Students** | **Institutions** | **Students** | **Institutions** | **Students** | **Institutions** | **Students** |
| Loan repayment, completion | 0% | 0% | 0% | 0% | 3% | 3% | 2% | 2% |
| Cohort default rate, completion | 0% | 0% | 0% | 0% | 0% | 2% | 0% | 2% |
| Completion, earnings | 0% | 0% | 20% | 8% | 1% | 0% | 1% | 1% |
| Cohort default rate, loan repayment | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% |
| Total | 0% | 0% | 20% | 8% | 5% | 6% | 4% | 5% |
| Total number failing | 0 | 0 | 2 | 328 | 12 | 6,121 | 14 | 6,449 |

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

# Notes

[1] Data from the Digest of Education Statistics 2020, table 326.10.

[2] See Opportunity America Working Group on For-Profit Colleges (2021) for a discussion of institutional growth rates as an indicator of quality.

[3] Michael Itzkowitz, "Price-to-Earnings Premium: A New Way of Measuring Return on Investment in Higher Ed," Third Way, April 1, 2020, https://www.thirdway.org/report/price-to-earnings-premium-a-new-way-of-measuring-return-on-investment-in-higher-ed.

[4] Among other problems, earnings may be underreported for workers relying on tips, such as cosmetology graduates (Kelchen 2020).

[5] More specifically, the adjustment operates as follows. If *nat_mean* is the national mean household income, and *zip_median* is the median household income from the college students' zip codes, an "adjusted" threshold centered at $15,000 is

*Adj_threshold* = $15,000 + $15,000 * *adj_factor* * (*zip_median* – *nat_mean*) / *nat_mean*

[6] Itzkowitz, "Price-to-Earnings Premium."

[7] "Official Cohort Default Rates for Schools," US Department of Education, Office of Federal Student Aid, last updated September 29, 2021, https://www2.ed.gov/offices/OSFAP/defaultmanagement/cdr.html.

[8] "Gainful Employment Information," US Department of Education, Office of Federal Student Aid, accessed February 7, 2022, https://studentaid.gov/data-center/school/ge.

# References


Ahlman, Lindsay, Debbie Cochrane, and Jessica Thompson. 2016. "A New Approach to College Accountability: Balancing Sanctions and Rewards to Improve Student Outcomes." Working paper. Oakland, CA: The Institute for College Access and Success.

Baum, Sandy, and Saul Schwartz. 2018. *Unaffordable Loans: When Should Schools Become Ineligible for Student Loan Programs?* Washington, DC: Urban Institute.

Blagg, Kristin, Erica Blom, Robert Kelchen, and Carina Chien. 2021. *The Feasibility of Program-Level Accountability in Higher Education*. Washington, DC: Urban Institute.

Blagg, Kristin, and Matthew Chingos. 2016. *Getting Risk Sharing Right: Creating Better Incentives for Colleges and Universities*. Washington, DC: Urban Institute.

Briones, Diego, and Sarah Turner. 2021. "Performance Measures and Post-Secondary Investments for Adult Students: Available 'Yardsticks' and the Challenges of Institutional Comparisons." Working paper. Charlottesville: University of Virginia.

Carnevale, Anthony P., Stephen J. Rose, and Ban Cheah. 2014. *The College Payoff: Education, Occupations, Lifetime Earnings*. Washington, DC: Georgetown University Center on Education and the Workforce.

Cellini Stephanie, Rajeev Darolia, and Lesley J. Turner. 2020. "Where Do Students Go When For-Profit Colleges Lose Federal Aid?" *American Economic Journal: Economic Policy* 12 (2): 46–83.

Cellini, Stephanie Riegg, and Claudia Goldin. 2014. "Does Federal Student Aid Raise Tuition? New Evidence on For-Profit Colleges." *American Economic Journal: Economic Policy* 6 (4): 174–206.

Cellini, Stephanie Riegg, and Nicholas Turner. 2019. "Gainfully Employed? Assessing the Employment and Earnings of For-Profit College Students Using Administrative Data." *Journal of Human Resources* 54 (2): 342–70. https://doi.org/10.3368/jhr.54.2.1016.8302R1.

Chou, Tiffany, Adam Looney, and Tara Watson. 2017. *A Risk-Sharing Proposal for Student Loans*. Washington, DC: Brookings Institution.

Cielinski, Anna, and Duy Pham. 2017. "Equity Measures in State Outcomes-Based Funding: Incentives for Public Colleges to Support Low-Income and Underprepared Students." Washington, DC: Center for Law and Social Policy, Center for Postsecondary and Economic Success.

Clotfelter, Charles T. 2012. *Measuring Colleges' Impact: Synopsis Paper*. Austin, TX: HCM Strategists.

Clotfelter, Charles T., Helen F. Ladd, Clara G. Muschkin, and Jacob L. Vigdor. 2013. "Success in Community College: Do Institutions Differ?" *Research in Higher Education* 54:804–24.

Deming, David J., and Kadeem Noray. 2020. "Earnings Dynamics, Changing Job Skills, and STEM Careers." *Quarterly Journal of Economics* 135 (4): 1969–2005. https://doi.org/10.1093/qje/qjaa021.

FSA (Office of Federal Student Aid). 2020. *Cohort Default Rate Guide*. Washington, DC: US Department of Education, FSA.

———. 2021. "Audits, Standards, Limitations, and Cohort Default Rates." In *2020–2021 Federal Student Aid Handbook, volume 2: School Eligibility and Operations*, edited by the FSA, 2-85–120. Washington, DC: US Department of Education, FSA.

GAO (Government Accountability Office). 2018. *Federal Student Loans: Actions Needed to Improve Oversight of Schools' Default Rates*. Washington, DC: GAO.

Hillman, Nicholas. 2016. "Designing and Assessing Risk-Sharing Models for Federal Student Aid." Working paper. Madison: University of Wisconsin–Madison, Wisconsin Center for the Advancement of Postsecondary Education.

Kantrowitz, Mark. 2013. *Consequences of the 90/10 Rule*. Las Vegas: Edvisors.

Kelchen, Robert. 2020. "Using Earnings Metrics for Accountability." Washington, DC: Higher Learning Advocates.

Klor de Alva, Jorge. 2022. "Accounting for Demographics, Selectivity, and Risk in Postcollege Earnings." In *Student Outcomes and Earnings in Higher Education Policy*, edited by Jason D. Delisle, 112–33. Washington, DC: American Enterprise Institute.

Klor de Alva, Jorge, and Mark Schneider. 2016. "A Risk-Sharing Model to Align Incentives and Improve Student Performance." Tempe, AZ: Nexus Research and Policy Center.

Lee, Vivien, and Adam Looney. 2019. "Understanding the 90/10 Rule: How Reliant Are Public, Private, and For-Profit Institutions on Federal Aid?" Washington, DC: Brookings Institution.

Minaya, Veronica, and Judith Scott-Clayton. 2020. "Labor Market Outcomes and Postsecondary Accountability: Are Imperfect Metrics Better Than None?" In *Productivity in Higher Education*, edited by Caroline Hoxby and Kevin Stange. Chicago: University of Chicago Press.

Opportunity America Working Group on For-Profit Colleges. 2021. *Accountability in Higher Education: For-Profit Colleges and Beyond*. Washington, DC: Opportunity America.

Speer, Jamin D., and Margaret Leighton. 2020. "Labor Market Returns to College Major Specificity." *European Economic Review* 128 (C). https://doi.org/10.1016/j.euroecorev.2020.103489.

Webber, Douglas A. 2018. **"**Risk-Sharing in Higher Education: A Policy Proposal to Improve Outcomes for Students and Reduce the Taxpayer Burden of Student Loan Debt." Bonn, Germany: Institute of Labor Economics.

# 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.

## Statement of Independence

The Urban Institute strives to meet the highest standards of integrity and quality in its research and analyses and in the evidence-based policy recommendations offered by its researchers and experts. We believe that operating consistent with the values of independence, rigor, and transparency is essential to maintaining those standards. As an organization, the Urban Institute does not take positions on issues, but it does empower and support its experts in sharing their own evidence-based views and policy recommendations that have been shaped by scholarship. Funders do not determine our research findings or the insights and recommendations of our experts. Urban scholars and experts are expected to be objective and follow the evidence wherever it may lead.


500 L'Enfant Plaza SW
Washington, DC 20024

*www.urban.org*