Postsecondary Equity & Economics Research Project
A joint initiative of
George Washington University,
Columbia University and Student Defense

Using a High School Earnings Benchmark to Measure College Student Success

Implications for Accountability and Equity

Stephanie Riegg Cellini
George Washington University and NBER
scellini@gwu.edu

Kathryn J. Blanchard
George Washington University
kjblanch@gwu.edu

JANUARY 2022

We thank seminar participants at GWU for helpful comments. We appreciate the support of the Arnold Ventures (Grant ID #20-044). The views expressed in this report are solely those of the authors.

In this report, we examine the design, performance, and implications of a high school earnings metric for accountability in higher education. This type of metric compares the earnings of former college students to the earnings of individuals with only a high school diploma. A version of this metric was recently proposed to assess student outcomes for short-term vocational programs that seek to access federal student aid.[1] One advantage of such a metric is the potential for a clear, simple, and intuitive threshold against which to measure the gains from postsecondary education, but there has been little research about the optimal design and implications of such a measure.

Earnings measures are central to conversations about accountability in higher education—particularly for short-term and vocational programs—and it is imperative that these measures accurately reflect student outcomes. Appropriate accountability metrics must be in place to ensure that students obtain a sufficient return on investment and are not taken advantage of by predatory institutions. At the same time, these metrics must ensure that institutions are not punished simply because of the demographics of the students they serve. This paper generates new evidence to inform debates over the use of earnings for higher education accountability and equity, with the hope of aiding the design of future measures.

We first describe the theory motivating a high school earnings metrics and discuss the challenges and choices involved in its design. We then develop two simple potential high school metrics based on Census data. The lower of our two benchmarks, at $25,000, is roughly two times the poverty line for an individual and aligns with the benchmark used by the Department of Education (DoED) in the College Scorecard in 2013-14. Today, the $25,000 benchmark is close to the earnings of young adults age 25-34 who did not complete high school.

We apply this benchmark to the most recent release of program-level earnings data for undergraduate programs in the Department of Education's Gainful Employment data.[2] Of 8,144 programs, just over half fail our low high school benchmark, with most failures accounted for by certificate programs (83%) and programs in for-profit colleges (76%). GE data on associate's and bachelor's degrees is limited to programs in for-profit colleges, but bachelor's degree programs fare well, with just 11% failing the low benchmark.

Student demographics do not appear to be driving failure of programs on a high school metric. Conditioning on institutional characteristics and demographics, our regression analyses show that programs in HBCUs are not disproportionately likely to fail high school earnings metrics and there are only negligible associations between the percentage of students of color and failure on high school metrics. Although not causal, our analyses suggest that sector, level, and institutional size are stronger predictors of failure than demographics, with for-profits, less-than-two-year, and larger institutions more likely to have programs that fail.

We suggest that future research assess differential effects by field of study, examine the distribution of students across failing programs, consider the outcomes of more recent cohorts of students, and account for programs that have may closed since the release of the GE data we use here. However, our findings suggest that adding a national or state-specific high school benchmark to debt-to-earnings metrics under Gainful Employment could potentially generate needed accountability for institutions and ensure value to students, without penalizing institutions for the students they serve.

BACKGROUND

In the upcoming negotiated rulemaking sessions, the Department of Education will undoubtedly look to strengthen accountability under Title IV of the Higher Education Act. Up for consideration will be whether to reinstate the 2014 Gainful Employment (GE) rule. The GE rule required nearly all programs in for-profit institutions and non-degree programs in other sectors to meet benchmark debt-to-earnings ratios. We compare program performance under GE to a proposed high school earnings metric below.

High school earnings metrics have been proposed previously, including for short-term vocational programs in recent attempts to reauthorize the Higher Education Act. For example, the College Affordability Act (H.R. 4674) put forward by House Democrats in 2019, proposed that—in addition to gainful employment and other requirements—programs between 300-600 clock hours would be required to show that the higher of the mean or median earnings of graduates are higher than the national average or (if justified) a state or local average for students with just a high school diploma to obtain loans. Similarly, Pell Grant eligibility for short-term public and nonprofit programs would need to meet "anticipated earnings" benchmarks agreed to by industry or sector partnerships, with the requirement that anticipated earnings be higher than local or national averages for individuals with only a high school diploma.

In the economics of education literature, researchers have proposed similar metrics for accountability. Matsudaira and Turner (2020) suggest holding all postsecondary programs accountable for one of two criteria – one based on graduates making progress on paying back their loans, and one based on a net earnings premium where most students must have earnings exceeding median earnings of the group with degrees one level lower. In the case of undergraduate programs, that comparison group would be high school graduates—the benchmark proposed here.

The Economic Justification for a High School Metric

The first economist to consider the economics of education was Gary Becker in his Nobel-prize winning work and his 1964 book, *Human Capital.* He argued that individuals decide whether or not to pursue additional education by comparing the lifetime benefits of that education—the largest of which is earnings—with the costs of that education. In the 1970s Mincer took this theory further, developing a model (now the famous "Mincer equation") that describes an individual's earnings as a function of schooling and work experience.

Many studies over the years have used the Mincer equation to estimate the average monetary return to an additional year of education. These types of estimates are commonly described as the "internal rate of return" and referred to simply as the "return" to education by economists. Most simply, this is just the increase in earnings or the "earnings gain" or "wage premium" attributed to college attendance. Comparing these earnings gains to the full cost of the investment in education - including both explicit monetary costs (e.g., tuition, interest) and implicit opportunity costs yields (e.g., the value of foregone earnings) yields the net present value of the education or a return on investment.

Mincer and others in this literature argue that—after controlling for work experience, age, and other confounding factors—earnings from employment reflect the value of a student's skills to employers in the labor market. The theory posits that an employer would not hire an individual if the value of an hour of that person's time were not at least as valuable as to them as the hourly wage they must pay her. Similarly, wages also reflect a person's own valuation of her time. In theory, an individual will only choose to accept a job if it pays an hourly wage at least as high as the value of an hour of their time spent in other ways, or they would not accept it. After investing in a year (or many years) of education, if the skills obtained are valued by employers, wages should rise. Although not a perfect measure, earnings are the most widely accepted measure of the benefits to education for economists, as they reflect the value of that person's time to society and indeed to themselves. They reflect the market value of a person's human capital.

Although Mincer's equation laid the groundwork for estimating the main benefits of education, without further adjustments the Mincer equation can only show correlation between earnings and education, and not causation. Therefore, the central concern in the vast literature estimating the earnings gains from college attendance in the many decades since Mincer, has been a focus on the extent to which estimates can credibly control for confounding factors and the selection that drives them.

The idea of selection is that students make choices about whether to attend college and which one to attend. Institutions also make choices of which students to accept in their admissions process. If, for example, higher-ability (or more wealthy, white, male, etc.) students end up in "better" (e.g., more selective) colleges, any raw differences in earnings after college may reflect these demographics and characteristics rather than the causal earnings gains generated by the college alone. Addressing this selection problem is a key concern in the literature and presents a challenge in accountability policy. I describe these challenges in the next section.

Earnings Metrics for Accountability

The theoretical measure of lifetime earnings gains are not practical in the context of accountability for two key reasons: timeliness and accuracy.

First, policymakers need a timely metric to protect students from low-performing programs. Lifetime earnings take many decades to be realized and measured accurately. A worker who does not attend college or take time out of the labor force, starts working at age 18, and retires at age 67 (the current age for full Social Security benefits for those born in 1960 or later) will work 49 years. For policy purposes, Social Security uses the highest 35 years of earnings to calculate benefits. Reasonable estimates of lifetime earnings are somewhere in the range of 35-50 years for most individuals and in the economics literature. Of course, policymakers cannot wait this long to assess student outcomes in the context of accountability.

Scott-Clayton and Minaya (2020) explore the correlation between earnings metrics measured in the short and longer-term with and without regression adjustment for lifetime earnings. They find that earnings measured two years after college completion without any adjustment are highly correlated with earnings after seven years. Their results suggest that earnings measured two-year post-college will be indicative of later earnings as well as lifetime gains. From a policy perspective, two years allows time for students to find new employment in their chosen field—while also allowing for timely consequences in an accountability framework. For shorter vocational programs, a shorter time period (e.g., 6 months or 1 year) might be reasonable and relevant assuming that students are more directly connected to the workforce after completion. In our data, we are constrained to use earnings measured three-years post-college completion. We contend that this time frame is sufficient for understanding student outcomes and generating timely feedback for policymakers.

A second problem facing policymakers is that calculation of the "gains" from education can be difficult to measure accurately. To measure gains, economists can employ instrumental variables estimation, regression discontinuity designs, or other quasi-experimental methods to generate a measure of gains free from selection. However, these methods are difficult to employ on a large scale for all students and schools in an accountability framework. A more feasible option to mitigate selection bias would be to compare a measure of the earnings a person has before attending to their earnings after. The idea is that all unobservable characteristics of students that do not change over time (e.g., race, gender, the quality of previous education) can be held constant, mitigating selection bias.

Obtaining "after" or post-college earnings is not a problem for government regulators, as the Department of Education has—in the past—partnered with the IRS and SSA to match students with post-college earnings data under Gainful Employment and the College Scorecard. However, in addition to post-college earnings, we also would ideally like a measure of pre-college earnings in order to accurately capture the after minus before earnings "gain" from college attendance.

We would ideally like to obtain earnings for a number of years before attending to see the trajectory of earnings changes due to college. With detailed data it is sometimes possible to see these pre-college trajectories for "nontraditional" (or "new traditional") students who enroll in college after several years in the labor market. However, it is nearly impossible for "traditional" students who enroll in college immediate following high school. These younger students may not work at all before attending college. Any earnings they report while in high school may be lower than the true value of their human capital for a number of reasons. For example, many high school students work just part-time after school or only at summer job, so annual earnings reported the IRS do not reflect the value of their full-time employment. Some students take low-paid or unpaid internships in order to "get a foot in the door" for a future career. Economists refer to these types of nonmonetary benefits from some jobs as compensating differentials. The idea is that the nonmonetary aspects of the job compensate for the lower wage. Still other students may work only informally or infrequently as babysitters, dog walker, or other odd jobs and fail to report those earnings to the IRS or SSA. For all of these reasons, economists typically omit any reported earnings of students before age 18.

One way to get around the problem of missing pre-college earnings, would be to assume a reasonable value of pre-college earnings for each student for whom we do not have "before" data. One could generate predictions using regression analysis to control for student background characteristics, demographics, location, etc. based on data from a large number of workers with no college education in the U.S. These could then be used as baselines for different groups of students (e.g., white non-Hispanic females with no children living in California). Predictions would be time consuming and might vary depending on which variables were included and which data sets used. Moreover, these estimates, while technically more accurate than using simpler approaches, would still be subject to various sources of error and potential biases from omitted variables, potentially raising doubts for those with limited statistical training. Such estimates could be problematic if different "predicted" earnings are used for different demographic groups, generating concerns around the inequity of applying different standards for different students and/or risk creating a culture of lower expectations for different groups. Finally, implementation and updating of these regression-based benchmarks could be challenging for policymakers. Interpretation of these types of estimates could be confusing for students.

A Simple Solution: High School Earnings Metric

A high school earnings metric may be a simple and reasonable approach to assessing the gains to education for undergraduate programs. With this approach, one would establish a benchmark "lower-bound" of earnings that we would expect the average student to have before college. This essentially substitutes for the "before" piece of the earnings gain or the "counterfactual" of the earnings that would be expected in the absence of the education. Then, one would consider the post-college earnings of students against this benchmark, asking whether post-college earnings met or exceeded the earnings for the average young high school graduate with no college education. Rather than rely on regression analyses, the comparison with "high school only" serves as a useful, if approximate, lower bar.

We investigate the implications of this type of benchmark "high school only" metric for accountability. Ideally, post-college earnings would be measured at the program-level for cohorts of students who exited the college in the same year. The highest of the median or mean earnings including both completers and (ideally) students who withdrew without completing would be computed and compared against the "high school only" benchmark.

Intuitively, if colleges are seeking taxpayer funds specifically for postsecondary education under Title IV of the Higher Education Act, there is a compelling argument makes that those programs should be able to show some value-added to earnings beyond secondary education. If institutions cannot show that their students have earnings higher than high school graduates with no college, then it could be argued that, in principle, they should not be offered funding through Higher

Education Act and the postsecondary education system more generally. These programs could seek funding elsewhere–for example, through workforce training programs. Moreover, many vocational education and training institutions not only exist, but seem to thrive in the U.S. without the benefit of Title IV funds. These institutions also charge lower tuition than similar programs and institutions accessing Title IV funds (Cellini and Goldin 2014).

Beyond ensuring taxpayer value, the larger goal of accountability policy is student protection. Students taking on federal student loans for postsecondary education deserve assurance that their earnings will increase beyond what they could expect if they had not attended. By definition, students who attend college (postsecondary education) should have already attained a high school diploma or GED (secondary education), so a benchmark based on high school or GED completion is intuitively appealing for students and easy to understand.

High school benchmark comparisons could be particularly useful for student decision-making. Higher education is commonly described by economists as exhibiting imperfect information where students—and particularly low-income students and those in underrepresented groups—may not be fully informed about the earnings gains and costs of various college options (e.g., Hoxby and Turner 2015). This lack of information is compounded by the fact that college education is an experience good, where the full value of a product (i.e., college) may not be known until after buying it (i.e., enrolling). Students may not be aware of the full costs, earnings gains, or employer's perceptions of the skills of graduates and may overestimate their ability to complete a degree. One goal of the College Scorecard website was to make information on student outcomes (including a high school benchmark) more easily available to prospective students to improve college choices, although it has had limited success in reaching lower-income students (Hurwitz and Smith 2017). The federal government has high-quality data on student earnings and costs that can and should be used to ensure that students can expect a minimal earnings boost from their education before taking on any amount of taxpayer-supported debt.

DEVELOPING METRICS

We begin by developing two potential national high school earnings benchmarks and then consider state-specific benchmarks.

For the national benchmarks, we draw on annual earnings estimates from the Current Population Survey's (CPS) Annual Social and Economic (ASEC) Supplement. The CPS is a joint effort between the Census Bureau and the Bureau of Labor Statistics (BLS). Data are collected annually from 60,000 households and estimates are considered the primary source of labor statistics for the U.S. population.

We show a number of potential national benchmarks based on the CPS and made publicly available in Table 1.

We consider two potential benchmarks based on this table in our analysis and classify them simply as: low and high. Our analysis focuses on two potential benchmarks that encompass a wide range for comparison, but other benchmarks falling between these could also be justified and applied. The metrics we choose for our initial exercise are static, but could be adjusted and updated annually to account for changing labor market conditions and data availability.

Key to our choice of benchmarks is the age range of the individuals on which the high school benchmark is based. The CPS reports data in two categories that are most useful and appropriate for recent college graduates, ages 18-24 and 25-34. Either of these age ranges could be justified for benchmarks, with the 18-24 range as a lower estimate and 25-34 as potentially more reasonable given that the average entering age of two-year college students is around 26 and earnings are measured several years later.

In contrast, using wages for all workers age 18 and older or age 25 and older (as typically reported by BLS), as shown in panel B of Table 1, would be problematic for college accountability. Since earnings increase with age, metrics based on groups that include older workers (e.g., above age 40) will be higher than what might be reasonably expected for recent college graduates 2-3 years after exit. We suggest that the estimates in panel A for younger groups are more appropriate for use as accountability benchmarks.

A second consideration is whether to base a benchmark on earnings for full-time full-year workers only or those with any reported earnings. Of course, full-time full-year workers have higher earnings for any given group, and the "any earnings" group will account for part-time workers and those who are unemployed or out of the labor force for part of the year. In theory, either group could be justified in developing a benchmark for accountability. Full-time full-year earnings—for a level of education below the college program under consideration—seems an appropriate goal for colleges to meet considering that full-time employment is also a goal of college-going (particularly for vocational programs under Gainful Employment) and would accurately reflect the counterfactual of what someone could earn without college. On the other hand, including part-time and part-year workers in a benchmark is more generous to colleges, allowing for unemployment and low hours.

Our most conservative "low" benchmark is $25,000 per year. It aligns with the benchmark used by the DoED in the first release of the College Scorecard in 2013-14. Justifying the use of this benchmark, the Scorecard notes, "The $25,000 threshold was chosen since it approximately corresponds to the median wage of workers age 25 to 34 with a high-school degree only." The $25,000 figure is simple, straightforward, and serves as a lower-bound relative to other earnings benchmarks.

We also choose this benchmark since it is still relevant today, but has a different interpretation: the median earnings of those as 25 to 34 who *did not even complete high school* was $25,536 in 2019, as highlighted in Table 1. Moreover, $25,000 roughly corresponds to 200% of the in 2019 (the year of our earnings data) for a single person living alone at $24,980. It also roughly corresponds 150% of the federal poverty line for households of two people at $25,365.

We propose $25,000 as a simple, intuitive lower bound for this analysis and for policy. We suggest that this benchmark—or a similar one adjusted annually—is the lowest justifiable benchmark for accountability. Intuitively, if colleges are typically enrolling high school graduates with the promise of higher earnings, it seems more than fair to expect that earnings after college should be above the average for high school dropouts in a similar age range. Moreover, the poverty line is already used in other student loan program benchmarks: Income Based Repayment plans typically offer zero monthly payments to borrowers with incomes below 150% of poverty, based on their household size.

For robustness, we consider other reasonable national benchmarks that are higher than $25,000. We believe a very strong case can be made to use the benchmark of $30,738 for young adults age 25-34 with any work experience (including part-time and part-year work) or $36,061 the median for this group with full-time work (see columns (4) and (2), respectively, of Table 1). To economize, we split the difference, and instead explore a $32,787 "high" benchmark between these two values which is also the mean for high school grads aged 18-24 working full-time in column (2). We generally prefer median earnings to means as they are less sensitive to outliers, but we use the mean value here as it is easily accessible in the CPS data tables and falls neatly between the two medians for age 25-34 that we prefer for policy more generally.

The Challenge of Location

Our high and low national benchmarks were chosen to generate values that might help policymakers assess the impact of similar measures in this range. One important drawback of using a national earnings benchmark is that some states and localities have higher or lower wages than others due to differences in local labor markets.

To address this issue, policymakers might consider drawing on state-level earnings data to create different high school benchmarks for each state. But, since many institutions and programs—and particularly online institutions—enroll students from around the United States, a more reasonable policy would be to allow institutions to appeal based on state or local earnings data, if the institution can show that its students are draw from and/or later employed in the local area.

The best state- and local-level data for earnings benchmarks comes from the Census Bureau's American Community Surveys (ACS), which produces earnings estimates for each year for geographic areas, such as states and smaller areas, such as Congressional Districts. We use the estimates from the Public Use Microdata Sample to construct state-level benchmarks for our analysis.

We conduct our analyses based on the median earnings for high school graduates aged 24-35 in the state. Appendix Table A1 reports these benchmarks. Our main analyses use the more conservative benchmark based on the broader group of individuals in the labor force (including those who work part-time and part-year or are unemployed and searching for work). As a robustness check, we show results based on full-time full-year workers in the Appendix Table A2. Notably, the ACS data align well with the national CPS data: across all states the median in the ACS data is $25,169 and the median of the 50 states' (and DC) medians is $24,918.

We view the ACS data as the most reliable publicly-available data on earnings for states and local areas, although there are other, less desirable, options. One option would be adjust the national averages based on other differences in average earnings for all workers across states or particular local areas. Another potential option policymakers might consider are state unemployment insurance (UI) data systems to develop benchmarks. However, UI records have several drawbacks relative to the ACS. State UI systems typically do not release aggregate earnings by education level or age, making it difficult to derive a "high school only" benchmark. Even if such benchmarks could be generated from state UI data systems, these systems do not cover all workers. Exclusions could be particularly problematic if certain educational programs have large numbers of students entering excluded fields. Exclusions differ by state, but typically exclude self-employed workers, federal workers, students working for schools, student nurses and interns in hospitals, real estate and insurance agents working on commission, individuals employed by relatives, and maritime workers, among other groups (Department of Labor 2020, p.7-9).

The Challenge of Timing

As previously noted, the timing of post-college earnings measures are important considerations. For shorter-term vocational programs, earnings could be measured 6 months to 1 year after attendance then compared to a high school benchmark to be most relevant for accountability. For longer-term programs, such as associate's and bachelor's degrees that may be less tied to a specific occupation, longer term earnings measures—such as 2-3 years after exit—may provide a better assessment of longer-term gains if findings a higher-skilled jobs might have larger search costs (e.g., a student might move or take time to explore different occupations). For timing choices, tradeoffs are involved: more consumer protection requires timely assessment but shorter post-college periods risk inaccuracy in assessing longer-term labor market outcomes.

DATA

We assess the performance of programs and institutions against a high school earnings metric and the correlates of failing such a measure, using the 2017 release of the 2015 debt year Gainful Employment (GE) Data. The GE earnings data contains 8,637 unique programs, 8,144 of which are undergraduate programs. These programs are spread across 2,616 institutions. Our sample includes non-degree undergraduate programs in all sectors, and associate's and bachelor's degree program in for-profit colleges (under the definition of gainful employment programs receiving Title IV aid under the Higher Education Act). Student outcomes are measured for graduates three years post-graduation and we use the highest of the mean or median earnings for comparison. We adjust earnings measures from 2016 to 2019 dollars.

METHODS

We begin by generating descriptive statistics and calculating the percentage of programs and type (sector, level, field, enrollment, etc.) that would pass or fail various high school earnings metrics. We next consider each metric in conjunction with debt-to-earnings measures proposed under the 2014 Gainful Employment Rule. We then assess the correlates of failure for both metrics and the combination. We pay close attention to institutional characteristics with the goal of assessing whether the high school earnings metric disproportionately affects institutions or programs that serve underrepresented groups. We assess the relationship between racial and ethnic composition and earnings metrics using data from IPEDS at the institution-level (program-level data on race and ethnicity is not available to our knowledge). We examine programs in HBCUs as well as institutions with a majority of Black and Hispanic students.

To better understand the role that different institutional, programmatic, and demographic characteristics play in passing or failing standards, we implement a simple linear probability model of the form:

$$Fail_{pis} = \beta_0 \boldsymbol{Sector}_i + \beta_1 \boldsymbol{Level}_i + \beta_2 \boldsymbol{Size}_i + \beta_3 \boldsymbol{RaceEth}_i + \beta_4 Female_i + \beta_5 Pell_i + \beta_6 HBCU_i + \beta_7 Online_i + \boldsymbol{d}_s + \varepsilon_{pis}$$

The outcome, *Fail,* takes on several different forms, but in each case the outcome of failure equals 1 and passing equals 0 for a program *(p)* in institution *(i),* and state *(s).* We run separate regressions for the following outcomes: failing each high school metric (low and high), failing GE, and failing either the earnings threshold or GE in combination.

Sector is a vector of indicators for for-profit, public, and nonprofit, where the omitted category is nonprofit. **Level** includes indicators for 4-year, 2-year, and less-than-2-year, with 4-year the omitted category. **Size** is based on bins of enrollment based on the distribution, 1-199 students, 200-999, 1000-4999, 5000+, with the smallest schools as the reference group. **RaceEth** includes separate variables for the share of students identifying as Black, the share of students identifying as Hispanic/Latino/Latinx, and the share identifying as other race/ethnicity. *HBCU* is indicator for Historically Black Colleges and Universities, *Female* the percentage of women, *Pell* is the percentage of Pell-eligible students, *Online* is an indicator variable that equals 1 if 50% or more students are enrolled exclusively online. Finally, $\boldsymbol{d}_s$ is a vector of fixed effects for each state and ε_{pis} is an error term. Standard errors are robust and clustered at the institution level to account for serial correlation.

Our approach can control for several key characteristics of programs that could influence the likelihood of failure that are not directly related to institutional quality and can therefore go further than comparisons of means in isolating the impact of college programs on student outcomes. Moreover, the state fixed effects can control for differences in earnings across states that are unrelated school quality (e.g., New York has higher average earnings for all individuals and West Virginia has lower). However, this simple OLS with fixed effects approach has important limitations. Notably, it cannot fully control for unobservable differences between programs and institutions (e.g., program length, instructor experience, etc.) that might drive differences, nor can it fully control for student selection on unobservables (e.g., students with lower ability selecting into certain programs). If, even after controlling for observable demographics and characteristics, students who have lower unobserved ability choose certain institutions—with otherwise equal quality—we might see lower earnings and less likelihood of passing high school benchmarks in the school with the lower ability students. We therefore consider these regressions only descriptive in nature and our results suggestive.

RESULTS

As shown Table 2, a sizable 55% of undergraduate GE programs would fail the lowest high school earnings benchmark of $25,000 in column (2). Put another way, 4,510 programs' graduates have post-college earning below the average for dropouts ages 25–34. The vast majority of these programs (3,753) are certificate programs and 62% of all certificate programs in the GE data fail.

In Table 3, we consider the characteristics of programs that fail the lowest threshold. Not surprisingly, certificate programs (in any sector) disproportionately fail the high school metric, accounting for 83% of failing programs, while comprising 75% of all programs in the data. But, breaking them down by sector in panel B, we see that certificate programs in for-profit college disproportionately fail and those in public institutions disproportionately pass. Looking only among for-profit programs in panel C, certificates and associate's degree programs fail in similar proportions to their composition, but the small number of bachelor's degrees again fare better.

To begin to explore the potential equity impacts of a high school earnings benchmark, we examine passage and failure of these metrics for HBCUs and institutions serving more than 50% Black, Hispanic, and Pell Grant eligible students in Table 4. All of the 24 certificate programs in HBCUs that are subject to GE (all HBCUs are public or nonprofit) fare well under the 2014 GE measure with none failing, but 10 of them (or 42%) would fail a $25,000 high school earnings benchmark. Similarly, programs in institutions where more than half of students are Black fare better on GE than on a high school benchmark. Among GE programs in these institutions 72% fail a high school benchmark, with even higher percentages failing among majority Hispanic institutions (86%), but lower percentages among majority Pell institutions (62%). These patterns could be driven by certain types of programs offered (e.g., certificates or fields of study) or institutions (e.g., for-profits) with poor outcomes that disproportionately recruit and enroll marginalized students.

To better understand these patterns, Table 5 presents the results of our regression analyses for the sample of certificate programs. The coefficients show the correlation between institution or program characteristics holding constant the other variables, indicating the independent contribution of the factor to passage or failure, although we note that these should not be interpreted causally. Columns (1) and (2) show results of regressions with outcomes of failing the "low" ($25,000) and "high" ($32,787) high school earnings metrics, respectively. Controlling for other characteristics, such as size and demographics, the regressions show that for-profits are significantly more likely to fail than the reference group, nonprofits, while programs in public institutions are less likely to fail. Programs in four-year institutions are less likely to have failing programs. Larger institutions are more likely to have failing programs at the low threshold, but coefficients on size are not significant for the higher threshold.

Demographic characteristics are shown in the lower rows of Table 5. There appears to be little correlation between the share of Black and Hispanic students and passing/failing the high school metric. While the share Black and Hispanic students are positively correlated with failure of both thresholds, the coefficients are extremely small. The coefficients suggest that—controlling for other factors—a one percentage point increase in the share of Black students increases the likelihood of the failing the lowest high school earnings threshold by just .0002 percentage points. Similarly, a one percentage point increase in the share of Hispanic students increases the likelihood of failing by .0001 percentage point. Scaling these up, even a 10 percentage point increase in the share of Black or Hispanic students (i.e., 20 percent of students to 30 percent) would only increase the likelihood of failing by .001 or .002 percentage points—an economically negligible magnitude. Calculating effects at the mean percent Black (19%) suggests marginal effects of 0.001% and percent Hispanic (19.5%) of 0.005%. Percent of Pell Grant recipients and percent female have similarly significant, yet extremely small effects.

For comparison, columns (3) and (4) run the same regressions, but using GE debt-to-earning metrics alone as the outcomes (either failing or in the warning zone), as in the 2014 GE Rule. For this rule, we see no significant correlation with for-profit status or level, but find that size is highly correlated with larger institutions faring worse on GE metrics. By demographics, results are similar, but percent other race is now small and negative while percent Hispanic is no longer significant. Importantly, HBCU status is negative and significant at the 10% level for GE metrics, indicating that HBCUs are less likely to fail GE debt-to-earnings metrics when controlling for other factors.

In columns (5) and (6) we combine the GE and high school metrics, counting as a failure one or the other. Correlations are nearly identical to the high school earnings metrics alone, suggesting that the high school metrics would generate greater accountability and drive failures in a combined measure.

Table 6 presents the results of similar regressions for associate's and bachelor's degree programs in for-profit colleges that are represented in the GE data. Relative to bachelor's degree programs and conditional on the covariates, associate's degree programs are more likely to fail the high school benchmarks, but are no more likely to fail GE (in col. 5), likely due to the low debt that students incur. Two-year institutions are slightly less likely to fail the high-threshold relative to four-year institutions, but no more/less likely to fail the lowest threshold, either alone or in combination with GE. Neither percent Black or Hispanic are significant at conventional levels for any threshold and coefficients remain small. Interestingly, percent other race/ethnicity is significantly negatively correlated with failing the "high" high school benchmark. Percent Pell Grant recipients in a school is again positively correlated with failure of all thresholds, but magnitudes remain small in columns (5) and (6).

RESULTS BY STATE

We break down failure rates by state and consider the use of state-level benchmarks in Table 7. Applying our benchmark of $25,000 to certificate programs in column (1), failure rates range from a low of 19% in Minnesota to a high of 90% in Delaware. When we allow for different thresholds for each state using CPS data in column (2), 22 states will see increases in failure rates, 12 stay the same, and 17 decrease.[19]

Columns (3) and (4) consider associate's and bachelor's degree programs in for-profit colleges included in the GE data. Note that nine states are missing data, presumably because no for-profit associate's or bachelor's degree programs met the necessary threshold for the number of students, or there are simply no Title IV for-profit degree programs in some low-population states.[20] Applying the standard $25,000 benchmark yields the full range of failure rates from 0-100%. If we apply state-specific benchmarks, 18 states would have more failures, 10 stay the same, and 14 decrease.

Ultimately, the state-level results are not surprising. They suggest that nearly an equal number of states and schools would benefit as would lose out with state-level benchmarks, rather than a national average or median. For policy, we suggest that schools in states with below-median earnings might be allowed to appeal based on state-level data. However, given how little our results changed with alternate benchmarks, we suspect that such an appeal would not be successful in many cases.

DISCUSSION AND POLICY RECOMMENDATIONS

A high school earnings threshold near $25,000 might be effectively layered on top of debt-to-earnings thresholds under the Obama Administration's 2014 Gainful Employment regulation. Intuitively, if colleges are seeking taxpayer funds specifically for postsecondary education under Title IV of the Higher Education Act, there is a compelling argument makes that those programs should be able to show value-added to earnings beyond secondary education. Students taking on debt for postsecondary education deserve assurance that their earnings will increase beyond what they could expect if they had not attended. Such a benchmark, while perhaps less precise than other proposed accountability measures, has an advantage in being simple to understand and easy to implement.

Our regression analyses suggest that sector, level, and institutional size are strong correlates of failure of a high school metric combined with debt-to-earnings rates under the 2014 GE rule. Programs in for-profit colleges, less-than-two-year institutions, and those larger institutions more likely to fail a high school benchmark than other types of programs and institutions. When controlling for these factors, status as an HBCU is uncorrelated with failure and percent Black and percent Hispanic have only negligible associations with failure of a combined high school and debt-to-earnings metric.

We suggest than any high school earnings metric should be adjusted annually to account for changing labor market conditions. Earnings benchmarks might also be allowed to vary by state, but the use of state-level benchmarks is unlikely to change overall failure rates nationwide: roughly similar numbers of states would see lower or higher failure rates. Of course, state-level benchmarks would benefit some states and programs more than others and could perhaps be productively applied in an appeals process. More research should be done to examine the implications of state-level benchmarks. We suggest that future research also assess differential effects by field of study, examine the distribution of students across failing programs, consider the outcomes of more recent cohorts of students, and account for programs that have may closed since the release of the GE data.

We recommend that policymakers reinstate the 2014 Gainful Employment debt-to-earnings rule and consider adding a layer of additional accountability through a high school earnings metric for the set of Gainful Employment programs. We contend that a benchmark based on the earnings of young high school graduates with no college could generate needed accountability for institutions and ensure value to students, without penalizing institutions for the students they serve. Although more research is needed, this type of benchmark is theoretically justifiable as a reasonable proxy for counterfactual or pre-college earnings by which to compare post-college earnings of students from IRS or SSA earnings. The metric is simple to implement and intuitive in the context of postsecondary education: we should expect colleges receiving aid under the Higher Education Act to generate earnings for their graduates that rise above those of high school students.

REFERENCES

Becker, G. S. (2009). *Human capital: A theoretical and empirical analysis, with special reference to education*. University of Chicago press.

Cellini, S. R. & Blanchard, K. J. (2021). "Quick College Credentials: Student Outcomes and Accountability Policy for Short-Term Programs." *Brookings* (blog). https://www.brookings.edu/research/quick-college-credentials-student-outcomes-and-accountability-policy-for-short-term-programs/.

Hoxby C. & Turner, S. (2015). What High-Achieving Low-Income Students Know About College. *American Economic Review*, 105(5): 514-517.

Hurwitz M. & Smith, J. (2018). Student Responsiveness to Earnings Data in the College Scorecard, *Economic Inquiry*, 56(2): 1220-1243.

Mincer, J. (1974). *Schooling, Experience, and Earnings*. Columbia University Press: New York.

Oreopoulos, P., & Petronijevic, U. (2013). Making college worth it: A review of research on the returns to higher education. *National Bureau of Economic Research Working Paper Series*, (w19053).

Matsudaira, J. & Turner L. (2020) "Towards a Framework for Accountability for Federal Financial Assistance Programs in Postsecondary Education." *Brookings* (blog) .

https://www.brookings.edu/research/towards-a-framework-for-accountability-for-federal-financial-assistance-programs-in-postsecondary-education/

Minaya, V. & Scott-Clayton, J. (2020). "3. Labor Market Outcomes and Postsecondary Accountability: Are Imperfect Metrics Better Than None?". *Productivity in Higher Education*, edited by Caroline M. Hoxby, Kevin Strange and Kevin Stange, Chicago: University of Chicago Press, pp. 67-104. https://doi.org/10.7208/9780226574615-005

Restrepo, L. & Turner, L. (2021). "Higher Education Performance and Accountability: Insights from a New Visualization Tool." *Brookings* (blog). https://www.brookings.edu/research/higher-education-performance-and-accountability-insights-from-a-new-visualization-tool/.

U.S. Department of Labor. (2020). "Comparison of State Unemployment Laws." <https://oui.doleta.gov/unemploy/pdf/uilawcompar/2020/coverage.pdf > Accessed on April 23, 2021.

Table 1. National High School Earnings Benchmarks, 2019$

		WORKED FULL-TIME YEAR-ROUND		ANY WORK EXPERIENCE	
		High School No Grad	High School Grad	High School No Grad	High School Grad
		(1)	(2)	(3)	(4)
A. Young Workers					
Age 18-24	Mean	34,854	32,787	12,899	21,594
Age 25-34	Mean	36,544	40,778	28,366	34,867
	Median	31,918	36,061	25,536	30,738
B. All Workers					
Over Age 18	Mean	40,858	47,833	29,082	39,371
Over Age 25	Mean	41,434	49,407	33,930	42,326
	Median	33,697	41,458	27,155	35,630

Source: U.S. Census Bureau, Current Population Survey, 2020 Annual Social and Economic Supplement (CPS ASEC).

https://www.census.gov/data/tables/time-series/demo/income-poverty/cps-pinc.html

Notes: "Worked Full-Time Year-Round" A year-round full- time worker is one who usually worked 35 hours or more per week for 50 weeks or more during the preceding calendar year. "Any Work Experience" Includes those persons who during the preceding calendar year did any work for pay or profit or worked without pay on a family- operated farm or business at any time during the year, on a part-time or full-time basis. Low, medium, and high benchmarks in bold. Numbers in bold indicate benchmarks used for data analysis.

Table 2. High School Earnings Metrics Applied to All Undergraduate Gainful Employment Programs

	Total	Failing Lowest	Passing Lowest
All Programs			
Number	8,143	4,510	3,633
Percentage	100%	55%	45%
Certificate Programs			
Number	6,081	3,753	2,328
Percentage	100%	62%	38%
Associates Programs*			
Number	1,464	694	770
Percentage	100%	47%	53%
Bachelors Programs*			
Number	598	63	535
Percentage	100%	11%	89%

Source: GE data

Notes: Counts and percentages of each program level that fail or pass the lowest high school metric. Certificate programs included are from for-profit, non-profit, and public institutions. *All associates and bachelors degree programs subject to GE are in for-profit institutions.

Table 3. Characteristics of GE Programs that Fail the Lowest High School Threshold

	Number	Percent Failing Low	Percent of Total Programs
A. Total failing	4,510	100%	
Certificate	3,753	83%	75%
Associates*	694	15%	18%
Bachelors*	63	1%	7%
B. Failing Certificate Programs			
Nonprofit	239	6%	6%
For-Profit	2,667	71%	54%
Public	847	23%	40%
C. Failing For-Profit Programs	3,424	76%	
Certificate	2,667	78%	75%
Associates*	694	20%	18%
Bachelors*	63	2%	7%

Source: GE data

Notes: * All associates and bachelors degree programs subject to GE are in for-profit institutions. Counts and percentages of failing programs of each level and sector.

Table 4. Minority-Serving Institutions' Program Outcomes

	Total	Fail Low High School	Fail GE
HBCU	24	10	0
		42%	0%
Over 50% Black	706	506	88
		72%	12%
Over 50% Hispanic	885	764	44
		86%	5%
Over 50% Pell	5,417	3344	654
		62%	12%

Source: GE data merged with IPEDS data measured at the institution-level.

Notes: HBCU is an indicator for Historically Black College and Universities.

Table 5. Correlates of Failure of High School Earnings Metrics and Gainful Employment Debt-to-Earnings, Certificate Programs

	High School Thresholds Only		GE Thresholds Only		High School or GE Thresholds	
	Fail Low	Fail High	Fail GE	Fail+Zone GE	Fail Low or Fail GE	Fail High or Fail GE
	(1)	(2)	(3)	(4)	(5)	(6)
Sector						
For-profit	0.136***	0.142***	0.013	0.025	0.133***	0.142***
	(0.033)	(0.030)	(0.014)	(0.033)	(0.033)	(0.030)
Public	-0.173***	-0.101***	-0.032**	-0.194***	-0.179***	-0.101***
	(0.039)	(0.036)	(0.014)	(0.033)	(0.038)	(0.036)
Level of Institution						
Less than 2-year	0.164***	0.136***	-0.025**	-0.010	0.162***	0.136***
	(0.026)	(0.024)	(0.011)	(0.025)	(0.026)	(0.024)
2-Year Institution	0.085***	0.070***	-0.013	-0.027	0.083***	0.070***
	(0.023)	(0.023)	(0.010)	(0.021)	(0.023)	(0.023)
Enrollment						
Enrollment 200-999	-0.011	0.018	-0.038***	-0.046***	-0.010	0.018
	(0.017)	(0.014)	(0.009)	(0.017)	(0.017)	(0.014)
Enrollment 1,000-4,999	0.005	0.011	-0.032**	0.004	0.007	0.011
	(0.024)	(0.022)	(0.013)	(0.025)	(0.024)	(0.022)
Enrollment 5,000+	-0.043	-0.010	-0.036***	-0.006	-0.042	-0.010
	(0.027)	(0.026)	(0.011)	(0.025)	(0.027)	(0.026)
Online Institution	-0.004	0.014	-0.000	-0.014	-0.004	0.014
	(0.016)	(0.013)	(0.008)	(0.015)	(0.016)	(0.013)
Race and Ethnicity						
Percent Black	0.002***	0.002***	0.001**	0.002***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
Percent Hispanic	0.001***	0.001***	-0	-0.001	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Percent Other	0.001	0.001*	-0.001**	-0.004***	0.001	0.001*
	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)
HBCU	-0.070	0.011	-0.049*	-0.109*	-0.068	0.011
	(0.090)	(0.089)	(0.028)	(0.060)	(0.090)	(0.089)
Percent Pell Recipients	0.002***	0.002***	0.000	0.001***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Percent Female	0.004***	0.002***	-0.000	0.001***	0.004***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	6,081	6,081	6,081	6,081	6,081	6,081
R-squared	0.318	0.218	0.066	0.179	0.318	0.218

Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$
Source: GE data merged with IPEDS data

Note: All regressions include state fixed effects. Omitted sector is nonprofit, omitted level is 4-year college, omitted enrollment is 1-199, omitted race is white. Online defined as more than 50% of students attending exclusively online. HBCU is an indicator for Historically Black Colleges and Universities. Fail low is defined as highest of mean/median program earnings below the lowest high school threshold. Fail mid is defined as highest of mean/median program earnings below the medium-low high school threshold. Robust standard errors clustered at the institution level.

Table 6. Correlates of Failure of High School Earnings Metrics and Gainful Employment Debt-to-Earnings, Associate's and Bachelor's Programs

	High School Thresholds Only		GE Thresholds Only		High School or GE Thresholds	
	Fail Low	Fail High	Fail GE	Fail+Zone GE	Fail Low or Fail GE	Fail High or Fail GE
	(1)	(2)	(3)	(4)	(5)	(6)
Associate's Degree	0.254***	0.317***	0.035	0.138***	0.140***	0.285***
	(0.028)	(0.030)	(0.032)	(0.039)	(0.032)	(0.031)
Level of Institution						
2-Year Institution	-0.002	-0.077**	-0.091**	-0.133***	-0	-0.077**
	(0.040)	(0.037)	(0.042)	(0.045)	(0.045)	(0.037)
Enrollment						
Enrollment 200-999	-0.076	0.091	0.103*	0.146*	0.003	0.092
	(0.054)	(0.055)	(0.060)	(0.076)	(0.060)	(0.056)
Enrollment 1,000-4,999	-0.115**	0.051	0.215***	0.213***	-0.021	0.062
	(0.058)	(0.060)	(0.063)	(0.078)	(0.065)	(0.061)
Enrollment 5,000+	-0.182***	-0.100	-0.051	-0.036	-0.199**	-0.111
	(0.066)	(0.075)	(0.084)	(0.102)	(0.078)	(0.077)
Online Institution	0.032	0.024	0.014	-0.025	-0.000	0.022
	(0.031)	(0.032)	(0.034)	(0.039)	(0.035)	(0.032)
Race and Ethnicity						
Percent Black	0.001	0.000	0.002	0.001	0.002	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Percent Hispanic	0.000	0.000	-0.001	-0.001	0.001	0.001
	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
Percent Other	-0.002	-0.006**	-0.006***	-0.010***	-0.005	-0.006***
	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)
Percent Pell Recipients	0.004***	0.004***	0.004***	0.005***	0.005***	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Percent Female	0.003***	0.002***	-0.001	0.000	0.001	0.002**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	2,062	2,062	2,062	2,062	2,062	2,062
R-squared	0.250	0.258	0.167	0.206	0.183	0.238

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Source: GE data merged with IPEDS data

Note: All regressions include state fixed effects. Omitted level is 4-year college, omitted enrollment is 1-199, omitted race is white. Online defined as more than 50% of students attending exclusively online. Fail low is defined as highest of mean/median program earnings below the lowest high school threshold. Fail mid is defined as highest of mean/median program earnings below the medium-low high school threshold. Robust standard errors clustered at the institution level.

Table 7. Failure Rates by State, Gainful Employment Data

State	Certificate Programs		Associates and Bachelors in For-Profits	
	Fail Low	Fail State Benchmark	Fail Low	Fail State Benchmark
AK	22%	71%	100%	60%
AL	59%	54%	40%	29%
AR	54%	54%	.	.
AZ	61%	64%	21%	22%
CA	75%	77%	28%	33%
CO	66%	74%	29%	39%
CT	43%	51%	0%	10%
DC	85%	85%	0%	0%
DE	90%	90%	.	.
FL	68%	64%	48%	37%
GA	65%	60%	31%	26%
HI	50%	75%	.	.
IA	54%	61%	39%	42%
ID	80%	81%	50%	50%
IL	68%	67%	28%	24%
IN	68%	67%	36%	31%
KS	45%	43%	30%	30%
KY	54%	49%	29%	29%
LA	72%	67%	35%	35%
MA	46%	60%	10%	29%
MD	62%	69%	13%	25%
ME	58%	58%	25%	25%
MI	65%	56%	.	.
MN	19%	28%	7%	25%
MO	56%	56%	49%	46%
MS	59%	57%	.	.
MT	40%	40%	.	.
NC	54%	46%	45%	9%
ND	31%	50%	.	.
NE	46%	54%	43%	43%
NH	68%	74%	24%	29%
NJ	65%	66%	17%	17%
NM	67%	67%	0%	0%
NV	81%	84%	17%	67%
NY	58%	57%	46%	37%
OH	51%	45%	62%	58%
OK	64%	64%	57%	57%
OR	64%	63%	24%	14%
PA	58%	58%	45%	46%
RI	67%	67%	.	.
SC	51%	41%	67%	50%
SD	41%	59%	8%	17%
TN	63%	52%	55%	33%
TX	63%	64%	44%	52%
UT	82%	90%	31%	54%
VA	60%	60%	42%	50%
VT	40%	40%	0%	25%
WA	55%	64%	14%	23%
WI	29%	33%	63%	75%
WV	62%	57%	55%	45%
WY	33%	44%	.	.

Notes: We use $25,000 as the low bar and show the percent of certificate programs in each state that fail the low bar, GE, or one or the other. The second section uses the median income of individuals with a high school degree aged 25-34 in the labor force (regardless of employment) in the 2018 ACS 5-year sample. We adjusted these earnings values to 2019$. Dots represent missing data (no programs meeting GE reporting thresholds in those states).

Appendix Table A1. State-level High School Earnings Benchmarks, CPS Data

State	Med. HS Earnings, in Labor Force	Med. HS Earnings, in Labor Force & Full Year
AK	27070	34607
AL	21657	26799
AR	23281	27634
AZ	25169	30103
CA	25587	30334
CO	27634	32484
CT	26611	32159
DC	21657	29364
DE	27085	32159
FL	23071	27070
GA	23583	27872
HI	29364	32484
IA	27267	32159
ID	24977	29251
IL	24564	29364
IN	25727	31087
KS	24918	29364
KY	23583	28817
LA	22751	28658
MA	28315	33558
MD	26799	32159
ME	24656	29777
MI	22083	28153
MN	28167	32484
MO	24797	29364
MS	20470	26218
MT	25587	29480
NC	22512	27085
ND	33558	37918
NE	26218	31294
NH	27872	32752
NJ	25727	31402
NM	23834	27872
NV	27070	32159
NY	24121	29364
OH	23583	29251
OK	25169	30230
OR	24564	29364
PA	25169	30705
RI	26799	32159
SC	21667	27070
SD	27267	32484
TN	21868	27085
TX	25587	29480
UT	27634	32159
VA	25587	31402
VT	26799	30909
WA	28153	32956
WI	26370	31418
WV	22338	28315
WY	30705	32752
US	25424	30403

Appendix Table A2. Failure Rates by State, Based on High School Graduates Working Full-Year

	Percent Failing Full Year Benchmark			
	GE Data		Scorecard Data	
State	Certificate	AA or BA in for-profits	Certificate	AA or BA (All)
AK	76%	80%	76%	43%
AL	64%	49%	50%	27%
AR	58%	.	60%	29%
AZ	75%	30%	68%	28%
CA	83%	51%	70%	35%
CO	80%	49%	66%	43%
CT	78%	30%	67%	32%
DC	85%	9%	73%	11%
DE	90%	.	83%	41%
FL	71%	58%	63%	25%
GA	71%	47%	53%	30%
HI	75%	.	86%	55%
IA	70%	56%	66%	29%
ID	84%	50%	55%	32%
IL	75%	50%	73%	27%
IN	78%	60%	62%	29%
KS	53%	50%	43%	22%
KY	63%	48%	71%	34%
LA	77%	35%	73%	36%
MA	71%	48%	62%	34%
MD	79%	63%	78%	35%
ME	79%	75%	61%	33%
MI	71%	.	70%	22%
MN	47%	43%	40%	28%
MO	66%	67%	54%	30%
MS	62%	.	61%	38%
MT	50%	.	73%	33%
NC	64%	55%	47%	31%
ND	50%	.	83%	34%
NE	68%	57%	69%	27%
NH	84%	47%	73%	32%
NJ	78%	43%	72%	31%
NM	67%	100%	71%	41%
NV	97%	100%	83%	39%
NY	69%	66%	56%	33%
OH	59%	84%	50%	29%
OK	71%	71%	68%	31%
OR	78%	43%	52%	36%
PA	67%	71%	63%	32%
RI	67%	.	50%	24%
SC	53%	67%	47%	27%
SD	65%	33%	60%	28%
TN	69%	67%	46%	27%
TX	70%	63%	68%	25%
UT	92%	81%	80%	36%
VA	71%	68%	65%	39%
VT	40%	25%	33%	33%
WA	76%	54%	59%	41%
WI	57%	100%	57%	29%
WV	73%	62%	60%	37%
WY	56%	.	67%	37%

ENDNOTES

1 We describe the proposal in the College Affordability Act (H.R. 4674) further below. https://edlabor.house.gov/the-college-affordability-act-facts

2 We describe these data and the 2014 Gainful Employment Rule in the next sections. https://studentaid.gov/data-center/school/ge

3 House Education and Labor Committee: https://edlabor.house.gov/the-college-affordability-act-facts. Bill: https://www.congress.gov/bill/116th-congress/house-bill/4674/text

4 For a nice summary see IZA (2016). https://wol.iza.org/uploads/articles/278/pdfs/estimating-return-to-schooling-using-mincer-equation.pdf

5 Social Security Administration website. https://www.ssa.gov/benefits/retirement/learn.html#h1

6 https://www.census.gov/programs-surveys/cps/technical-documentation/methodology.html

7 The average age at college entry of two-year colleges students is 26 and the average age for four-year college students is 23. Assuming college takes 2-4 years minimum, the average ages at completion would be about 27-30 and earnings are measured 2-3 post-exit. Estimates based on authors' calculations of the 2017 Digest of Education Statistics report.

8 https://collegescorecard.ed.gov/assets/FullDataDocumentation.pdf

9 HHS website 2021.

10 The poverty threshold for individuals living alone was $12,490 and $16,910 for two-person households for the 48 states and the District of Columbia (HHS website 2021).

11 We could not track down medians for the 18-24 age group in the publicly available CPS data tables we used, but one could calculate them using CPS microdata.

12 Data is from the Public Use Microdata Sample here: https://data.census.gov/mdat/#/. We thank Lesley Turner for sharing the downloaded and cleaned file for the 2018 vintage data. We adjust for inflation to 2019 dollars to match the year of our earnings data.

13 Seasonal workers interviewed in their off season are not included in the labor force.

14 The median of the states' median earnings in 2019 dollars is $25,169.

15 Graduate programs in for-profit institutions are included in GE, but we do not consider them here, as a high school earnings benchmark would be unreasonably low for graduate students.

16 We calculate percent Black, percent Hispanic, and percent women using enrollment numbers in each category and total enrollment at that institution, divide them and multiply by 100. We calculate the percent online variable similarly and convert it to a dummy variable equal to 1 if more than 50% of students are enrolled completely online.

17 State benchmarks are reported in Appendix Table 1A. Our main results use the benchmarks based on all individuals in the labor force in column (1). Additional state-level results using full-year workers are included in Appendix Table 2A with benchmarks listed in column (2) of Appendix Table 1A.

18 The missing states are AR, DE, HI, MI, MS, MT, ND, RI, WY.

19 State benchmarks are reported in Appendix Table 1A. Our main results use the benchmarks based on all individuals in the labor force in column (1). Additional state-level results using full-year workers are included in Appendix Table 2A with benchmarks listed in column (2) of Appendix Table 1A.

20 The missing states are AR, DE, HI, MI, MS, MT, ND, RI, WY.