URBAN INSTITUTE · ELEVATE · THE · DEBATE

Unintended Consequences of an Earnings-Based Accountability Test for Master's Degree Programs

An Essay for the Learning Curve by Cody L. Christensen

January 2024 (corrected January 11, 2024)

Students and families are concerned about rising college costs and growing student debt levels, which has prompted policymakers to begin holding postsecondary programs accountable for their students' future earnings. Early accountability efforts focused only on certificate programs and degree programs offered at proprietary institutions,[1] but policymakers are beginning to consider accountability measures for master's degree programs.[2]

But some stakeholders have raised concerns that an earnings-based accountability test might disproportionately penalize master's degree programs serving larger concentrations of students from disadvantaged gender and racial and ethnic backgrounds.[3] To the extent this accountability test would affect programs offering little value to students, penalization could be a good outcome.[4] But to the extent that lower performance on an earnings-based accountability test reflects factors such as occupational segregation or labor market discrimination against women and people of color, it could unfairly penalize these groups and the programs that serve them.

[1] For example, the original gainful employment rule, enacted in 2012, introduced a debt-to-earnings test for all certificate programs and degree programs at proprietary institutions. See Program Integrity: Gainful Employment—Debt Measures, 76 Fed. Reg. 34386 (Jun. 13, 2011).

[2] For example, in 2023, Senator John Cornyn (R-TX) introduced the Streamlining Accountability and Value in Education for Students Act, which would introduce an earnings-based accountability test for all master's degree programs. Programs that fail the earnings test in two out of three consecutive years would lose access to federal student loans. See Streamlining Accountability and Value in Education for Students Act, S. 1971, 118th Cong. (2023).

[3] For example, Matsudaira and Turner (2020) discuss how the legacy of racial segregation and discrimination at historically Black colleges and universities (HBCUs) may partially explain why many master's programs at these institutions are predicted to fail an earnings-based accountability test. See Jordan D. Matsudaira and Lesley J. Turner, *Towards a Framework for Accountability for Federal Financial Assistance Programs in Postsecondary Education* (Washington, DC: Brookings Institution, 2020).

[4] This, of course, assumes that students attending lower-quality programs can switch to higher-quality programs. These higher-quality programs could be at the same institution or at another nearby institution. Switching programs and colleges, however, can be complicated. At the undergraduate level, many students attending colleges that close do not enroll at a different institution. For more, see Rachel Burns, Ellen Bryer, Kelsey Heckert, Dustin Weeden, and Lynneah Brown, *A Dream Derailed? Investigating the Causal Effects of College Closures on Student Outcomes* (Boulder, CO: State Higher Education Executive Officers Association, 2023).

In this essay, I estimate how an earnings-based accountability test, which I call the "net earnings test" (NET), would affect master's degree programs and whether failing this test is correlated with the types of students the master's degree program serves. I find that approximately 70 percent of master's degree programs lead to positive returns and pass the NET. The remaining share would fail the test because those programs lead to negative returns for their typical students. These programs include some of the most common master's degree programs, such as clinical social work, marriage and family therapy, mental health counseling, and community health counseling.[5]

Master's degree programs that lead to negative returns disproportionately enroll larger shares of female students and Black students. Programs serving the largest concentrations of Black, nonwhite Hispanic, and female students have the lowest returns, on average.

These results imply that an earnings-based accountability test could disproportionately affect master's degree programs that serve larger shares of historically disadvantaged students. From a policy standpoint, some stakeholders would consider this an unintended consequence. A well-targeted accountably policy would affect programs based solely on the program's quality, not for the types or characteristics of students they serve. An earnings-based accountability test, though, cannot distinguish between programs that are truly of low quality from programs of fine overall quality but that struggle to produce strong earnings outcomes because of the students they serve and the greater challenges they face in the labor market.

Policy approaches such as including program-specific exemptions in the accountability test could mitigate this issue but could allow other low-quality programs to skirt accountability and undermine the policy's intended goal. In short, policymakers will face complicated trade-offs when designing and implementing accountability policies for master's degree programs. The collection of more detailed data about master's degree programs could reduce this challenge.

Defining the Earnings-Based Accountability Test

Policymakers could theoretically use various measures for accountability purposes (e.g., a program's price, loan repayment rate, or cohort loan default rate), but I focus on an earnings-based accountability test because achieving higher earnings is the most common reason students express for pursuing postsecondary education.[6]

---

[5] These are specific programs within the fields of "health and medical administrative services" and "mental and social health services and allied professions," which are two of the most common fields of study among all master's degrees.

[6] See "Why Higher Ed?" Gallup, accessed December 8, 2023, https://news.gallup.com/reports/226457/why-higher-ed.aspx. An additional reason that earnings are the basis for the accountability test is because program-level earnings information is readily available in the US Department of Education's College Scorecard data. Program-level information on master's degree program prices, for example, are not available.

I use a NET similar to the one proposed by Matsudaira and Turner (2020).[7] Specifically, I construct an earnings premium (EP) for every US master's degree program (with available data) using this equation:

$$EP = (Typical\ Earnings) - (Counterfactual\ Earnings) - (Out\ of\ Pocket\ Costs) \qquad (1)$$

Box 1 provides an example of how a program's EP is calculated. In short, the EP measures how much more a typical graduate from a master's degree program earns relative to a typical student with only a bachelor's degree in the same broad field of study and living in the same state, after accounting for the out-of-pocket expenses students pay to attend the master's degree program.

Using the EP as an earnings-based accountability metric has several advantages. It accounts for the counterfactual earnings that students would have likely experienced in their states' local labor market had they never pursued graduate school, allowing for a better determination about whether students are better off for having gone to graduate school. Second, the EP considers the high costs students pay to attend graduate school.[8]

A master's degree program with a positive EP implies that its graduates earn more than the typical bachelor degree holder in the same broad field of study living in the same state, after considering the cost of the master's program. Programs with negative EPs imply the opposite. Appendix A provides additional details on the data I use to construct program-level EPs. For this study, I consider programs with positive EPs as "passing" the NET and programs with negative EPs as "failing" the NET.[9]

---

[7] Matsudaira and Turner, *Towards a Framework for Accountability for Federal Financial Assistance Programs*. There are several differences between the earnings premium I propose here and the "net earnings premium" Matsudaira and Turner propose. For example, Matsudaira and Turner adjust program-level earnings using the program completion rate using deidentified institution-level information from Matsudaira from the original iteration of the College Scorecard. Second, they use Scorecard data from only the 2014–15 and 2015–16 school years, whereas I incorporate additional years of College Scorecard data. Lastly, I use observed median three-year earnings data (as reported in the College Scorecard), whereas Matsudaira and Turner estimate each program's median three-year earnings using one-year earnings data. They do this because three-year program-level earnings data were not reported or available in the College Scorecard at the time of their study. In the end, Matsudaira and Turner's metric uses earnings data for 940,000 master's degree completers, whereas I use earnings data covering more than 3 million completers.

[8] A typical master's degree program costs students around $11,000 in annual out-of-pocket tuition expenses after all grants and scholarships are applied. The EP accounts for those expenses when determining whether a student is better off for having attended graduate school. Author's calculations using 2019–20 National Postsecondary Student Aid Study, PowerStats.

[9] This study does not explore potential penalties for programs that fail the NET. But policymakers may wish to limit or restrict failing programs from receiving federal financial aid. For example, recent legislation introduced by Senator Cornyn (R-TX) would prohibit master's degree programs that fail an earnings-based accountability test in two out of three consecutive years from receiving student loans. For a detailed discussion of Senator Cornyn's bill, see Jason Delisle and Jason Cohn, "An Earnings Test for Master's Degrees: Identifying Programs at Risk of Failing a Proposed Rule for Federal Loans" (Washington, DC: Urban Institute, 2023). The earnings test in that bill differs slightly from the earnings test I use in this study. Notably, the Cornyn bill does not account for out-of-pocket expenses in its earnings test. Further, the Cornyn bill does not compare the earnings of master's degree holders with specific bachelor's degree holders in the same broad field in the same state. Instead, the Cornyn bill compares the earnings of master's degree holders with the average earnings of all bachelor's degree holders in the state.

BOX 1

Example Earnings Premium Calculation

This box explains how a program's EP is computed, using the master's of social work (MSW) program at the University of North Carolina at Chapel Hill (UNC) to illustrate this calculation.

- **Typical earnings.** Graduates of the UNC MSW program have a median income of $43,240 three years after they exit the program (all values are in 2016 nominal dollars, converted using the Consumer Price Index).
- **Counterfactual earnings.** Individuals ages 25 to 34 living in North Carolina who are not enrolled in college with a bachelor's degree in education and public service fields (see appendix C for a list of programs within this category) had a median income of $33,080.
- **Out-of-pocket costs.** The median amount of Stafford and Grad PLUS loans disbursed to graduates of the UNC MSW program is $47,400. Amortized over 25 years, this annual expense is $3,200.

A program's EP is calculated via the following formula:

$$EP = (Typical\ Earnings) - (Counterfactual\ Earnings) - (Out\ of\ Pocket\ Costs) \qquad \text{(Bx1)}$$

Plugging in the values above, the EP for the UNC MSW program is

$$EP = \$43{,}240 - \$33{,}080 - \$3{,}200 = \ \$6{,}960 \qquad \text{(Bx2)}$$

Because the EP is greater than $0, the program passes the NET. An EP is calculated for every master's degree program for which data are available. See appendix A for technical details.

Thirty Percent of Master's Programs Fail the Net Earnings Test

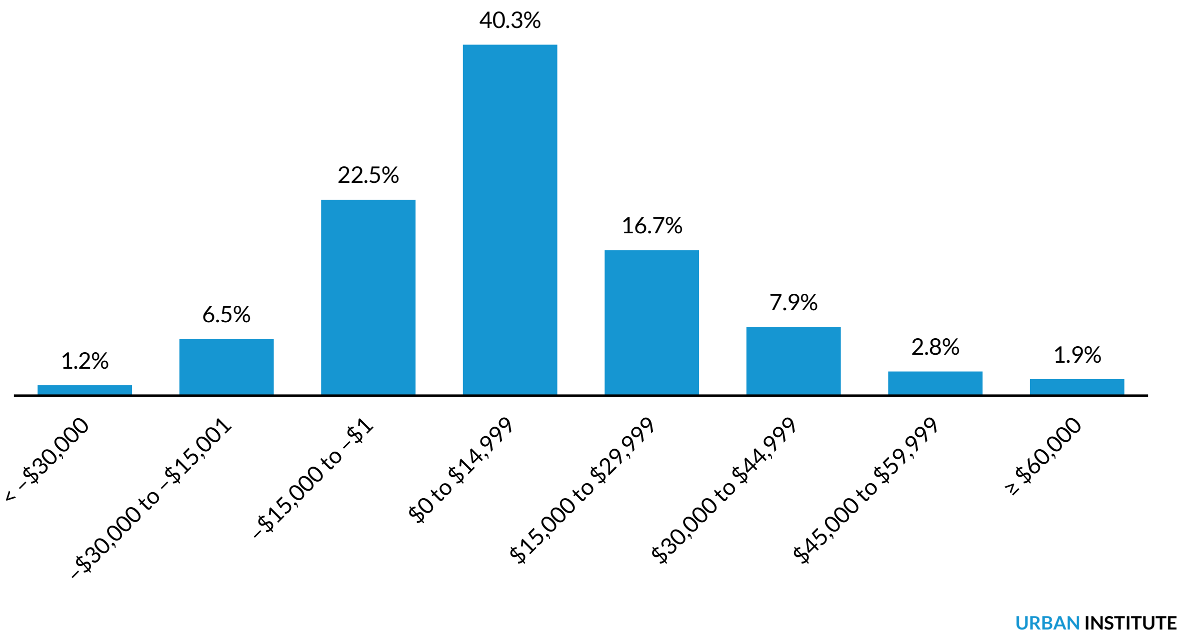
Figure 1 displays the EP distribution for nearly all US master's degree programs with available earnings data (*N* = 12,174 programs). Between 2014 and 2019, 3.2 million students graduated from these programs. Appendix A provides additional details on the sample of programs and students.

The average EP for all master's degree programs is $9,400.[10] Approximately 70 percent of master's degree programs have a positive EP, meaning they would pass the NET. The remaining 30 percent of master's degree programs would fail the NET because they have a negative EP.

[10] This is the unweighted program average in constant 2016 dollars. When weighting programs by program size (i.e., the number of completers in the program), the average earnings premium rises to $15,800.

FIGURE 1

Distribution of Programs, by Earnings Premium



Source: Author's calculations using data from the US Department of Education and the US Census Bureau.

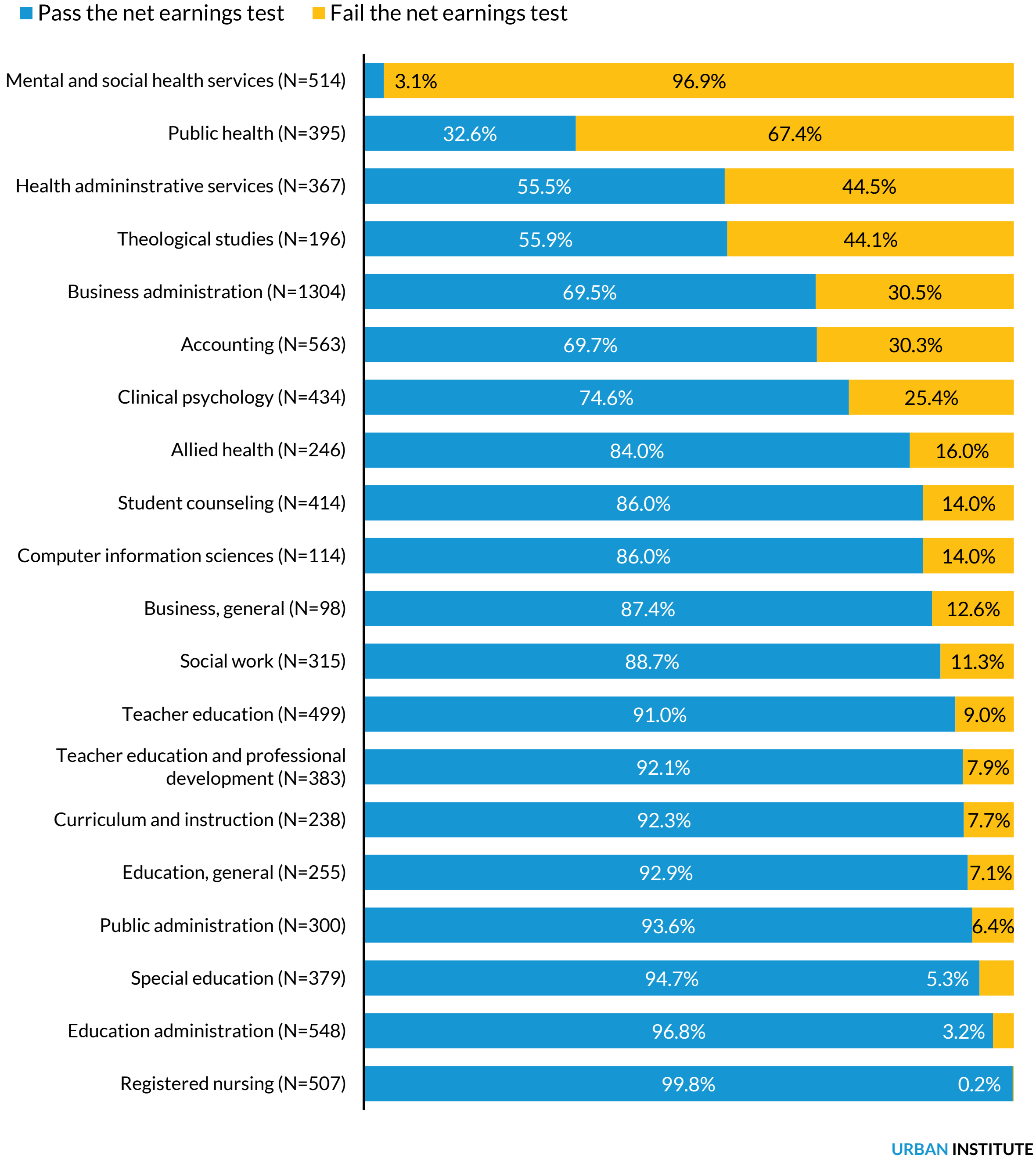
Notes: *N* = 12,174 programs. All dollar values are in constant 2016 dollars, adjusted using the Consumer Price Index. See the text and appendix A for details on the earnings premium measure. Appendix figure B.1 presents the share of students enrolled in programs affected by the net earnings test.

Certain fields of study pass the NET at higher rates than others. Figure 2 focuses on the NET pass rates for the 20 largest master's degree programs nationwide. Many of these programs would pass the NET. For example, business administration programs, the most common master's degree program, have an overall NET pass rate of roughly 70 percent, implying that 913 of the 1,304 business administration master's degree programs would pass the NET, while the remaining 391 programs would fail.

Some of the largest master's degree programs have remarkably high pass rates. For example, nearly 100 percent of registered nursing master's programs (*N* = 507 programs) and 97 percent of educational administration master's degree programs (*N* = 548 programs) pass the NET. Other large master's degree programs perform much worse. For example, just 3 percent of mental and social health services master's degree programs (*N* = 514 programs) would pass the NET.

FIGURE 2

Pass and Fail Rates for the 20 Largest Master's Degree Programs Nationwide



Source: Author's calculations using data from the US Department of Education and the US Census Bureau.
Notes: NET = net earnings test. Programs are defined at the four-digit Classification of Instructional Program code level. The 20 largest programs were determined by measuring the total number of students that graduated from the program between 2014 and 2019. *N* sizes refer to all unique programs that operated for at least one year between 2014 and 2019 with available earnings data. See the text and appendix A for details on the NET and earnings premium measures. Appendix figure B.2 presents the share of students enrolled in programs affected by the NET.

Appendix B presents the shares of students in programs that are affected by the NET.[11] Appendix figure B.1 shows that around 80 percent of master's degree completers between 2014 and 2019 graduated from programs that passed the NET, while the remaining share came from programs that failed the NET. Appendix figure B.2 presents the share of students affected in the 20 largest master's degree programs.[12]

Programs with Larger Shares of Female Students and Black Students Are More Likely to Fail

Programs that fail the NET have notable differences in the types of students they serve. Figure 3 shows that programs that fail the NET have larger shares of female students (66 percent) relative to programs that pass the NET (61 percent), on average. Similarly, programs that fail the NET have smaller shares of white students and larger shares of Black students relative to passing programs.[13]

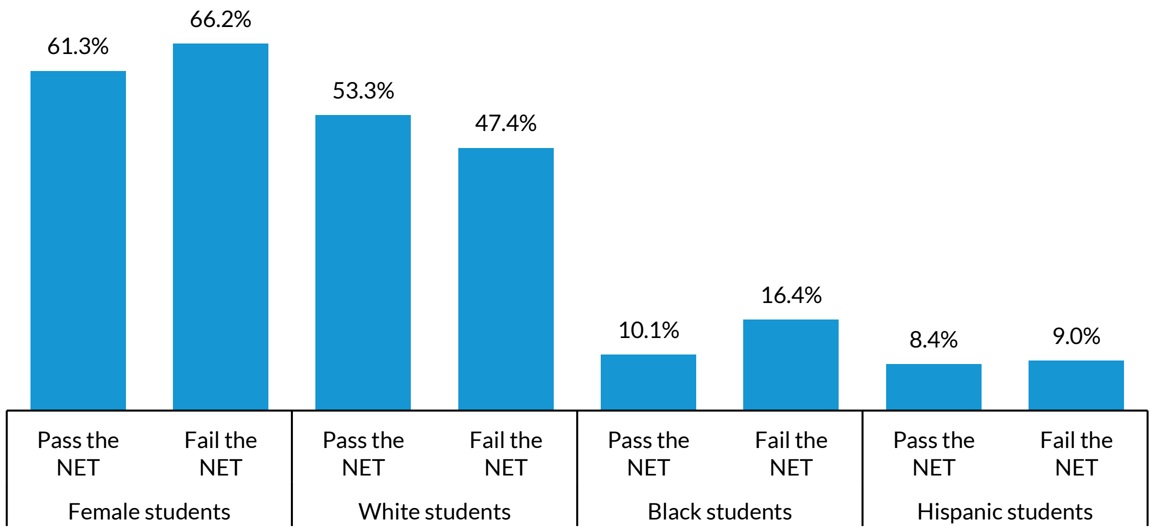
[11] Appendix table B.1 shows that master's degree programs in STEM (science, technology, engineering, and mathematics) and education fields have high NET pass rates. On the other end of the spectrum, more than half of all law, theology, and liberal arts master's degree programs fail the NET. Appendix figure B.3 shows that the average earnings premium varies widely across programs. Engineering, computer information technology, math, and health professions have some of the highest earnings premiums, whereas programs in culinary services, communication technologies, and performing arts fields have the lowest earnings premiums, on average. Appendix figure B.4 disaggregates earnings premiums by institutional control. Master's degree programs at public and private institutions usually have positive EPs (though theology programs at public institutions are a large exception). For-profit institutions offer programs with more of a mixture of positive and negative EPs. The lower earnings premium is driven by the higher out-of-pocket costs that students usually pay at for-profit master's degree programs.

[12] For example, in the student analysis in appendix figure B.2, 470,513 students graduated from business administration programs between 2014 and 2019 that pass the NET, while the remaining 117,628 business administration master's degree holders graduated from programs that fail the NET. Among the 20 most common master's degrees programs, mental and social health services stands alone as the only program where nearly all graduates attended programs that fail the NET.

[13] Appendix table B.2 reports descriptive differences in program characteristics for the other race or ethnicity categories (Asian and "other race"). These results indicate that failing programs have slightly fewer shares of Asian students and students of other races, on average. But these differences are not statistically distinguishable from zero.

FIGURE 3

Weighted Average Characteristics of Program Completers



Source: Author's calculations using data from the US Department of Education and the US Census Bureau.
Notes: NET = net earnings test. This figure reports the weighted average characteristics of program completers by whether the program passes the NET. Programs are weighted by the total number of completers in the program between 2014 and 2019. See the text and appendix A for details on the NET.

Beyond the pass-fail dichotomy, the concentration of Black, Hispanic, and female students in a master's degree program is correlated with program earnings premiums. Figure 4 divides all master's programs (*N* = 12,174) into 10 evenly sized deciles (each with roughly 1,217 programs in it) based on the share of students in the program that are female, white, Black, and Hispanic.

Programs with larger shares of female students, Black students, and Hispanic students have lower average EPs, whereas there is little association between the EP and the proportion of white students in the program. Although they are few in number, nearly all master's degree programs with substantial concentrations of Black students have low or negative EPs.[14]

This correlation means that programs at historically Black colleges and universities, which enroll larger concentrations of Black students, could be disproportionately affected by the NET. This disparity bears out in the data: the average EP of programs offered at HBCUs is $2,300, roughly four times lower than the average of programs all other institutions ($9,500). Programs at HBCUs are 25 percent more likely to fail the NET relative to programs at all other institutions.[15]

[14] When programs are grouped into 20 evenly sized quantiles, rather than the deciles used for figure 4, the 600 programs in the top quantile of Black students have a negative EP, on average.

[15] Forty percent of master's degree programs at HBCUs have a negative EP, meaning they would fail the NET. For comparison, only 30 percent of programs at all other institutions would fail the NET. That is, the fail rate at HBCUs is 25 percent higher than the fail rate at all other institutions.

FIGURE 4

Deciles of Earnings Premiums, by Student Demographic Characteristics

URBAN INSTITUTE

Source: Author's calculations using data from the US Department of Education and the US Census Bureau.
Notes: This figure groups the full sample of programs (*N* = 12,174) into 10 evenly sized deciles based on the share of students in the program that have a given demographic characteristic. The four panels, respectively, group programs into deciles by the share of female students, white students, Black students, and Hispanic students in the program. Each decile has roughly 1,217 programs.

Of course, there are many other inputs (e.g., the program's field of study, local labor market conditions, and institutional spending on instruction) that may explain why programs with larger concentrations of female, Hispanic, and Black students have lower earnings.[16] In Appendix D, I conduct a simple, descriptive analysis to examine whether the negative relationship between earnings outcomes and student characteristics within a program disappears when comparing programs that are identical on these other observable factors. The analysis suggests that these other characteristics do explain some of the gap in earnings outcomes, but they do not explain all of it. Said differently, earnings gaps persist across programs with different racial and ethnic and gender characteristics after accounting for other observable program characteristics.

Policy Implications

These findings may complicate efforts to enact accountability policies for master's degree programs because some low-quality programs—as measured by average earnings premium—may actually provide their students exceptional value and education experiences, but they would get labeled as "low quality" because the earnings outcomes of their students are lower because of additional challenges that their students face in the labor market.

How might policymakers distinguish between programs that are *legitimately* of low quality—such as those that mismanage resources, provide students with poor classroom instruction, and offer lackluster career support services—from programs that get *labeled* as being low quality simply because their students come from historically disadvantaged backgrounds and experience greater challenges in finding and securing high-paying jobs the labor market?

Below, I outline three ways policymakers could modify the NET to reduce the number of programs that are affected by the test, with the goal of assisting programs that serve large concentrations of students from historically disadvantaged populations. Each of these policy levers is imperfect and involves trade-offs that compromise different aspects of the NET.

Include Program-Level Carve-Outs to the NET

One approach would create carve-outs in the NET for certain types of programs or institutions. Exempted programs might include ones that propel many of their students into occupations that provide large social value but typically offer lower earnings. Another idea could exempt all programs at certain types of institutions, such as HBCUs or tribal colleges. These institutions have experienced a history of discrimination and continue to serve larger concentrations of disadvantaged students today.

This idea entails multiple trade-offs. Exemptions would lower the risk that policymakers inadvertently sanction programs of fine academic quality that fail the NET simply because they serve

[16] For example, master's programs in STEM, which have larger concentrations of men, may have higher earnings outcomes than master's programs in public service, which have higher concentrations of women, simply because STEM programs typically lead to jobs with higher earnings than public service programs. If that is the case, the negative relationship between program earnings and the gender characteristics of students in a program should disappear when considering only master's degrees within the same field of study.

larger concentrations of students from disadvantaged populations, but exemptions could also give a free pass to programs at exempted institutions that are mismanaging resources and offering poor instruction.

Vary the EP Thresholds Used in the NET Based on Student Characteristics

Policymakers could also vary the threshold for passing the NET based on the share of students with certain sociodemographic characteristics within a program. Unlike the prior policy lever, which categorically exempts certain programs (or institutions) from the NET, this idea would hold all programs accountable but would judge them differently based on student characteristics.

As a concrete example, policymakers could give programs a $5,000 “bonus” added to the program’s typical earnings (the earnings measure used to compute a program’s EP) if the program serves an above-average share of low-income students.[17] Of course, one challenge is that institutions (or programs) at risk of sanctions could game this metric by targeting their recruitment efforts at low-income students, thereby receiving a lower EP threshold to be judged against and ultimately avoiding penalties from the accountability test. Additionally, some stakeholders may critique this policy as giving certain types of institutions an undeserved free pass. Proponents of this view would argue that all programs, regardless of the types and backgrounds of students they serve, should be held to the same high standards.

Implement the NET Evenhandedly, Coupled with Targeted Investments

A third idea could apply the NET evenly to all master’s degree programs but do it in conjunction with a large investment in institutions that have experienced a legacy of discrimination, which could include targeted investments in HBCUs, tribal colleges, and other minority-serving institutions. In theory, these institutions can use this investment to improve instructional quality, facilities, and student support services, all of which may be important inputs when determining students’ earnings outcomes.

But this policy still has limitations. First, it is unclear how large or how frequent such an investment should be. Second, the targeted investments could be somewhat arbitrary. For example, two similar programs at different colleges that serve equal concentrations of Black students could receive different funding simply because one program is at an HBCU while the other is not. Third, it is unclear whether the targeted investment will be used in ways that enhance student outcomes. Institutions that receive the investment could spend it on things that do little to improve student outcomes, thereby failing to raise the program’s EP.

Lastly, one more potential benefit from an evenhanded NET is that it could lead institutions to lower tuition prices for their master’s degree programs, which could reduce student debt. Recall that median debt of program graduates, amortized over 25 years, is the proxy for out-of-pocket costs used in the NET calculation (box 1). Thus, increases in student earnings *or* reductions in student debt both increase

---

[17] This idea could be extended to student demographic characteristics, such as gender and race or ethnicity categories, though it is unclear whether such a policy would be constitutional. Setting that issue aside, giving master’s programs that serve the highest decile of Black students (i.e., where 30 percent or more of program graduates are Black) a $5,000 bonus in “typical earnings” (box 1) would reduce the number of failing programs in that decile by 15 percent. ($5,000 is approximately one-fifth of a standard deviation in typical earnings.)

a program's EP. Arguably, institutions have more control over the latter, making this a potentially viable path for institutions to consider if they are at risk of sanctions for failing the NET.

How much would programs need to reduce the debt of their students to have meaningfully large increases in their EPs? If all programs reduced their median debt by $10,000, which is approximately 25 percent of the average debt level of master's degree completers, that would reduce annual out-of-pocket expenses by an average of $700 per program.[18] All else equal, this reduction would result in approximately 160 master's programs that would have failed the NET absent the $10,000 median debt reduction to now pass the NET.[19] This analysis suggests that lowering prices (and thereby lowering student debt) can improve EPs, but price reductions would have to be sizeable to flip programs from failing the NET to passing the NET.

In the end, policymakers do not have the data to distinguish between programs that are labeled as low quality because they offer their students little educational value relative to programs that are labeled as low quality because they serve larger shares of students from disadvantaged backgrounds. To address this challenge, researchers and policymakers should capture additional information on program quality (e.g., educational outcomes, instruction quality outcomes, and career and student support services) to better distinguish whether programs with low earnings outcomes are attributable to factors that are under the program's control or to broad societal factors—namely, the possibility of racial and gender discrimination—that are outside an individual program's ability to address.

Appendix A. Constructing the Earnings Premium and Analytic Sample

This appendix describes how I constructed the earnings premium measure and how I selected the analytic sample of programs.

Computing Program EPs

The earnings premium measure compares the earnings of master's degree completers with the typical earnings of bachelor's degree holders in similar fields, after netting out the additional costs students pay to attend their master's degree program. Positive EP values imply the master's degree program leads to positive returns for its students, on average, whereas negative values imply the opposite. I consider programs with positive EPs (or EPs equal to $0) as "passing" the NET and programs with negative EPs as "failing" the NET.

[18] The average master's degree costs $11,000 in out-of-pocket tuition expenses, and master's degree students typically graduate with $40,000 in total student debt. Author's calculations using 2019–20 National Postsecondary Student Aid Study, PowerStats table akhfzg and College Scorecard data. See box 1 for a definition of out-of-pocket costs. The analysis regarding a $10,000 reduction in median debt assumes that median program debt is reduced by $10,000 for all programs or, for programs with median debt less than $10,000, median debt is reduced to $0.

[19] This is approximately a 1.4 percentage-point decline in the share of programs that fail the NET. One caveat to reducing program prices, of course, is that it could lead to reductions in institutional resources, which could negatively affect program quality. Said differently, increases in a program's EP from cutting tuition prices could be offset by declines in a program's EP through lower graduate earnings.

The method to construct the EP follows a similar method to the net earnings premium measure proposed by Matsudaira and Turner (2020). In this study, the EP is constructed from three components: typical earnings, counterfactual earnings, and out-of-pocket costs. I describe each component below.

TYPICAL EARNINGS

I define typical earnings as the median earnings of master's degree completers three years after program exit. I compute this measure using five pooled cohorts of program-level College Scorecard data (i.e., from the following pooled cohorts: 2014–15 and 2015–16, 2015–16 and 2016–17, 2016–17 and 2017–18, 2017–18 and 2018–19, and 2018–19 and 2019–20). A program's median three-year earnings measure is a weighted average of the five years of data, where each program's earnings outcome is weighted based on the average number of completers in the pooled cohort in the given year. Before averaging, I adjust all earnings measures to constant 2016 dollars using the Consumer Price Index (CPI). Programs that do not report any median earnings metrics are dropped from the sample. Later in this appendix, I describe how I impute median three-year earnings for programs with missing data.

Median earnings data used to derive each program's EP are based only on the earnings outcomes of program graduates who received federal Title IV aid. This means the earnings of students who did not receive Title IV aid are not included in a program's median earnings data. This limitation stems from the way postsecondary program data are collected and reported. Without better data, I rely on these metrics as proxies for the earnings outcomes for all graduates.

COUNTERFACTUAL EARNINGS

Counterfactual earnings are the median earnings of bachelor's degree holders (who are not currently enrolled in college) ages 25 to 34 in the same broad field of study living in the same state where the master's degree program is located. I construct this measure using data from the American Community Survey (ACS). I begin by pooling ACS data from 2014 to 2019. I limit the sample to individuals who have only a bachelor's degree ages 25 to 34 (inclusive). Students who are currently enrolled in college (at the graduate or undergraduate level) are dropped. Individuals with top-coded incomes are replaced to missing. I convert income to constant 2016 dollars using the CPI. Appendix C shows how I link master's degree programs in the College Scorecard (which are defined by Classification of Instructional Programs, or CIP, codes) to bachelor's degree major categories in the ACS. I compute the average income for individuals in each broad field of study for each state. For example, the counterfactual earnings for students pursuing a master's degree program in legal professions and studies at an institution in Kansas is the median individual earnings of bachelor's degree holders ages 25 to 34 living in Kansas who are not currently enrolled in college with a bachelor's degree in any of the following ACS major fields: law, social sciences, or business.

OUT-OF-POCKET COSTS

Out-of-pocket costs are operationalized as the median Stafford and Grad PLUS loan debt disbursed at the institution to Title IV students in the program, amortized over 25 years with a nominal 4.5 percent annual interest rate. I would prefer to measure the actual out-of-pocket costs that students spend on tuition, fees, textbooks, and other education-related expenses for their program, taking the median amount of that value and amortizing it over 25 years. But this information is not disaggregated at the program level for master's degree programs. That is why I instead use median Stafford and Grad PLUS loan amounts. Although these values may capture not enough (or too much) of the actual educational expenses that master's degree students pay to obtain their degree, these data are the best available information we have on program-level master's degree prices.

Constructing the Analytic Sample

The final sample contains 12,174 unique master's degree programs, from which 3.2 million students graduated between 2014–15 and 2019–20. This section describes how I selected this sample.

I begin by pooling five panels of College Scorecard data between 2014–15 and 2019–20. I adjust all earnings variables to constant 2016 dollars using the CPI. I collapse the panel to construct a single cross-section that is unique by college-program-award level. I keep only programs classified as master's degree programs in the College Scorecard data. I exclude master's degree programs that had no completers between 2014 and 2019. After these restrictions, I am left with 32,125 unique master's degree programs that had at least one graduate in one or more years between 2014–15 and 2019–20. Nearly 4.2 million master's degree students graduated from these programs between 2014–15 and 2019–20.

Next, I drop programs that do not have any median earnings data for any earnings metric reported in the College Scorecard data. Specifically, programs with missing data for one-, two-, three-, or four-year median earnings for all years are dropped. Programs that report one or more of those earnings measures during one or more years are kept in the sample. Programs that reported only one-, two-, or four-year earnings are adjusted to three-year earnings using the process described below. I exclude programs without any earnings data because I cannot construct their EP, which is the key variable in the analysis. This removes 58.8 percent of master's degree programs, which enroll 18.4 percent of master's degree completers. This restriction reduces the sample of programs to 13,248, which collectively had 3.4 million completers between 2014 and 2019.

I impose a few additional sample restrictions to make the sample representative of typical US master's degree programs. First, I drop a small number of master's degree programs offered at colleges in US territories (*N* = 150 programs), offered at community colleges (*N* = 3 programs), and in a trades field (CIP codes 46, 47, 48, and 49) (*N* = 7 programs). I also drop 119 programs in geographies that have no bachelor's degree holders within the same geography, meaning I have no earnings data to compare them with to generate the earnings premium. Lastly, I drop a small number of programs offered at colleges that do not appear in the Integrated Postsecondary Education Data System (IPEDS) completers data (*N* = 34 programs) or the IPEDS finance data (*N* = 271 programs). I made these restrictions so sample sizes are constant in the analysis.

Ultimately, I am left with 12,174 unique master's degree programs, which had 3.2 million completers between 2014–15 and 2019–20. Conditional on earnings data, this sample represents 92 percent of all master's degree programs, and these programs have 96 percent of master's degree graduates. I merge these data with IPEDS data to obtain demographic information about students who complete their programs. Importantly, this means student characteristics within programs (e.g., race, ethnicity, and gender) are based on program completers, not program enrollees. I use data on completers rather than enrollees because demographic information about program enrollees is not available in IPEDS.

Imputing Earnings for Programs with Missing Three-Year Earnings Outcomes

This section describes the method for imputing median three-year earnings data for the programs with missing three-year earnings. First, programs that report no median earnings data during any year are dropped from the sample. Of the remaining programs, 63 percent report three-year median earnings data for one or more years. If a program reports only a single year of median three-year earnings data between 2014 and 2019, I use that value for the program's typical earnings metric in equation 1. If the same program reports three-year earnings data in two or more years of the Scorecard data between 2014 and 2019, I average those values together, weighting by the average number of completers in the program in a given year. (All earnings values are adjusted to constant 2016 dollars using the CPI.)

Conditional on having earnings data, 37 percent of programs do not report three-year median earnings data in any year but do report median earnings data for cohorts one, two, or four years after program exit in one or more years of the College Scorecard data. I impute missing three-year median earnings data for programs by projecting forward one-year and two-year median earnings data out to the third year. Similarly, I project the four-year median earnings data backward to the third year. This way, all earnings data are measured from the same point three years after cohort exit.

The exact process for imputing missing three-year earnings data is as follows. First, for programs with missing three-year median earnings data and nonmissing two-year median earnings data, I adjust the program's two-year median earnings to use in place of the missing three-year median earnings. If the same program reports two or more years of two-year earnings data, I average the values together (adjusting for inflation and weighting by the number of program completers) before projecting the value forward. I project the earnings data forward using ACS data to estimate the average annual earnings growth rate for bachelor's degree holders ages 25 to 34 in the same broad field of study and living in the same state and apply the average growth rate to upwardly adjust the two-year earnings by one year.[20] For example, bachelor's degree holders living in Arizona in the arts and humanities field ages 25 to 34 have an average annual income growth rate of 3.5 percent. For master's degree programs in Arizona in

[20] I use a similar process for programs with missing two-year median earnings and missing three-year median earnings but with nonmissing four-year median earnings. I deduct the average one-year earnings growth rate such that the four-year median earnings adjust the earnings data backward to the third year. Lastly, for programs with missing two-, three-, and four-year median earnings but with nonmissing one-year median earnings, I upwardly adjust the one-year median earnings by projecting forward two years of the field's average income growth rate, such that the median one-year earnings are projected forward to the third year.

arts and humanities fields (see appendix C for the crosswalk) that report only two-year median earnings data, I upwardly adjust the two-year median earnings data using the following equation:

$$\text{Proxy 3-year Median Earnings} = (1 + 0.035) \times [\text{Median 2-year Earnings}] \tag{1}$$

In total, I impute missing earnings data for 4,545 programs (37 percent of all programs). Table A.1 presents a balance table in program characteristics across programs with observed and imputed three-year median earnings data. Although similar on some characteristics, programs with imputed three-year median earnings data have smaller shares of female students and Black students and larger shares of students from the "other race" category. When replicating the main analysis above using only programs with nonimputed earnings data, results are qualitatively similar.

TABLE A.1

Balance Test between Programs with Imputed and Nonimputed Earnings

	Programs with observed 3-year median earnings	Programs with imputed 3-year median earnings	Difference	Standard error
Female	0.666	0.625	0.041**	0.004
White	0.567	0.559	0.008	0.005
Black	0.130	0.106	0.024**	0.003
Hispanic	0.086	0.083	0.003	0.002
Asian	0.049	0.047	0.002	0.001
Other race	0.168	0.205	-0.036**	0.003
Observations	7,629	4,545		

Source: Author's calculations using data from the US Department of Education and US Census Bureau.

Notes: Values in this table represent the proportion of students in programs. * $p < 0.05$, ** $p < 0.01$

Appendix B. Characteristics of Master's Degree Programs, by Field of Study

TABLE B.1

Share of Students and Programs That Pass the NET, by Broad Field of Study

CIP group	*N* all programs (1)	*N* failing programs (2)	Share of failing programs (3)	*N* all students (4)	*N* students in failing programs (5)	Share of students in failing programs (6)
Business	2,226	683	30.7%	879,184	174,464	19.8%
Consumer and pub. service	1,359	384	28.3%	325,648	52,941	16.3%
Law	68	38	55.9%	21,995	8,116	36.9%
Health	2,203	944	42.9%	558,491	169,994	30.4%
Liberal arts	1,833	1,028	56.1%	328,312	170,127	51.8%
STEM	1,411	272	19.3%	395,088	51,310	13.0%
Education	2,872	229	8.0%	669,388	38,783	5.8%
Theology	202	105	52.0%	51,471	30,054	58.4%
Total	12,174	3,683	30.3%	3,229,575	695,787	21.5%

Source: Author's calculations using data from the US Department of Education and the US Census Bureau.

Notes: CIP = Classification of Instructional Programs; NET = net earnings test. Programs are placed into CIP groups using the crosswalk in appendix C. Column 1 reports the total number of unique programs with available earnings data and at least one completer from 2014 to 2019. Column 2 reports the number of programs in column 1 that failed the NET. Column 3 is a ratio of the values reported in column 2 and column 1. Column 4 reports the total number of students that completed a master's degree program between 2014 and 2019. Column 5 reports the number of students in column 4 that graduated from a program that failed the NET. Column 6 reports the ratio of the values reported in column 5 and column 4. See the text and appendix A for details on the NET and sample construction.

TABLE B.2

Differences in Student Characteristics between Programs That Pass and Fail the NET

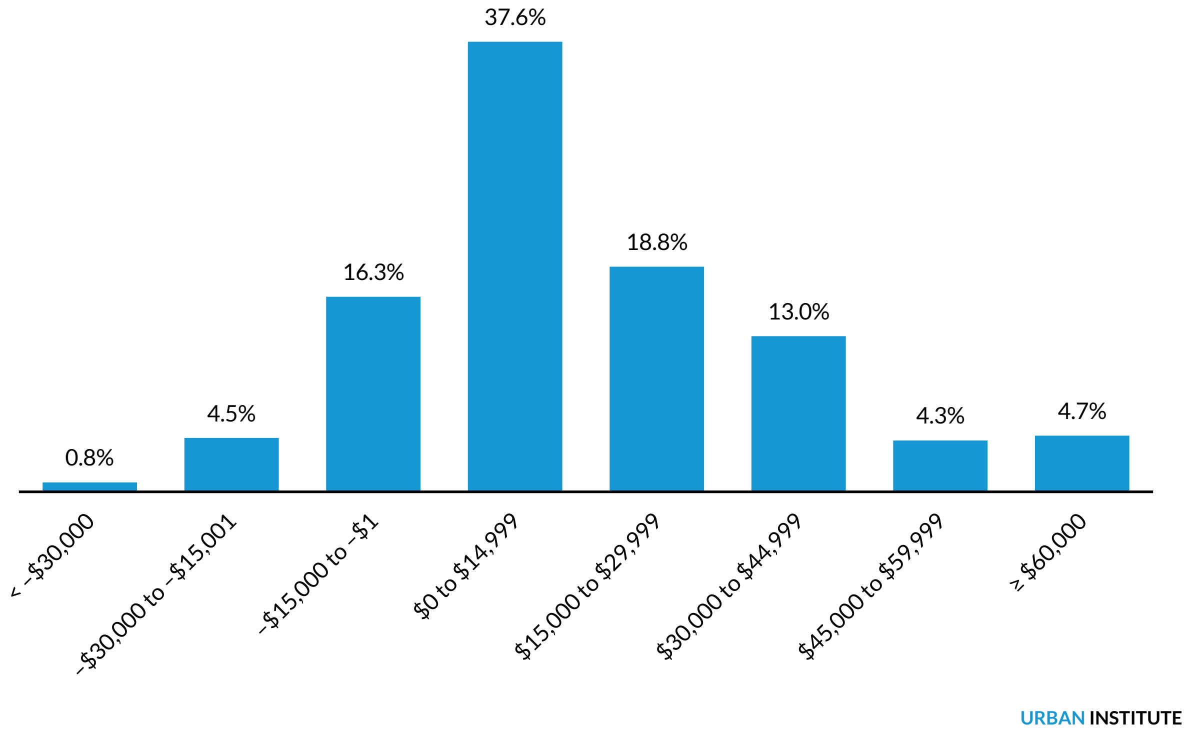
Program characteristic (1)	All programs (2)	Programs that pass the NET (3)	Programs that fail the NET (4)	Difference (5)	*p* value (6)
% Female	62.34	61.32	66.16	4.85	<0.001
% White	51.99	53.25	47.39	-5.86	<0.001
% Black	11.47	10.11	16.42	6.32	<0.001
% Hispanic	8.52	8.39	8.98	0.58	0.186
% Asian	5.78	5.89	5.37	-0.51	0.103
% Other race	22.25	22.36	21.84	-0.52	0.668
Observations	12,174	8,491	3,683		

Source: Author's calculations using data from the US Department of Education and the US Census Bureau.

Notes: NET = net earnings test. Program characteristics (column 1) are the characteristics of program completers only. Column 2 reports the average characteristics of all programs, weighted by program size. In columns 3 through 6, the values in each row are estimated in a separate regression of the variable listed in column 1 on a binary indicator equal to 1 if the program failed the NET. Specifically, column 3 is the regression constant, column 5 is the estimated slope coefficient, column 4 is the sum of columns 3 and 5, and column 6 is the *p* value associated with the slope coefficient. In columns 3 through 6, observations are weighted by the number of total completers in the program between 2014 and 2019. See the text and appendix A for a description of the NET. The *p* value in column 6 is associated with a two-sided hypothesis test, where the null hypothesis is that the coefficient in column 5 is equal to 0. Robust standard errors are used in all regressions. Standard errors are clustered at the institution level.

FIGURE B.1

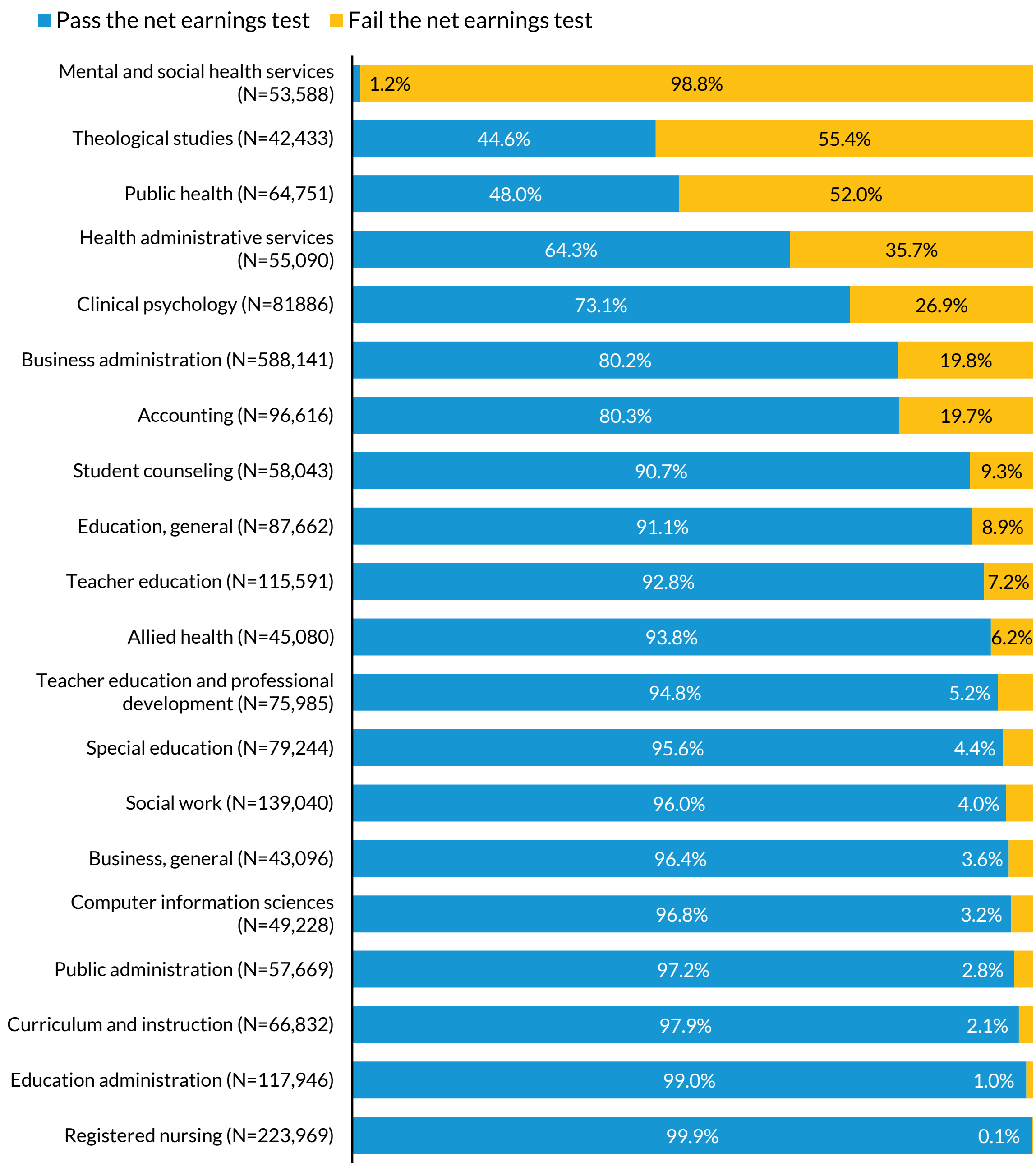
Percentage of Students in Programs, by Earnings Premium Bin



Source: Author's calculations using data from the US Department of Education and the US Census Bureau.
Notes: *N* = 12,174 programs. All dollar values are in constant 2016 dollars, adjusted using Consumer Price Index. See the text and appendix A for details on the earnings premium measure.

FIGURE B.2

Proportion of Students in the 20 Largest Master's Programs, by NET Pass or Fail

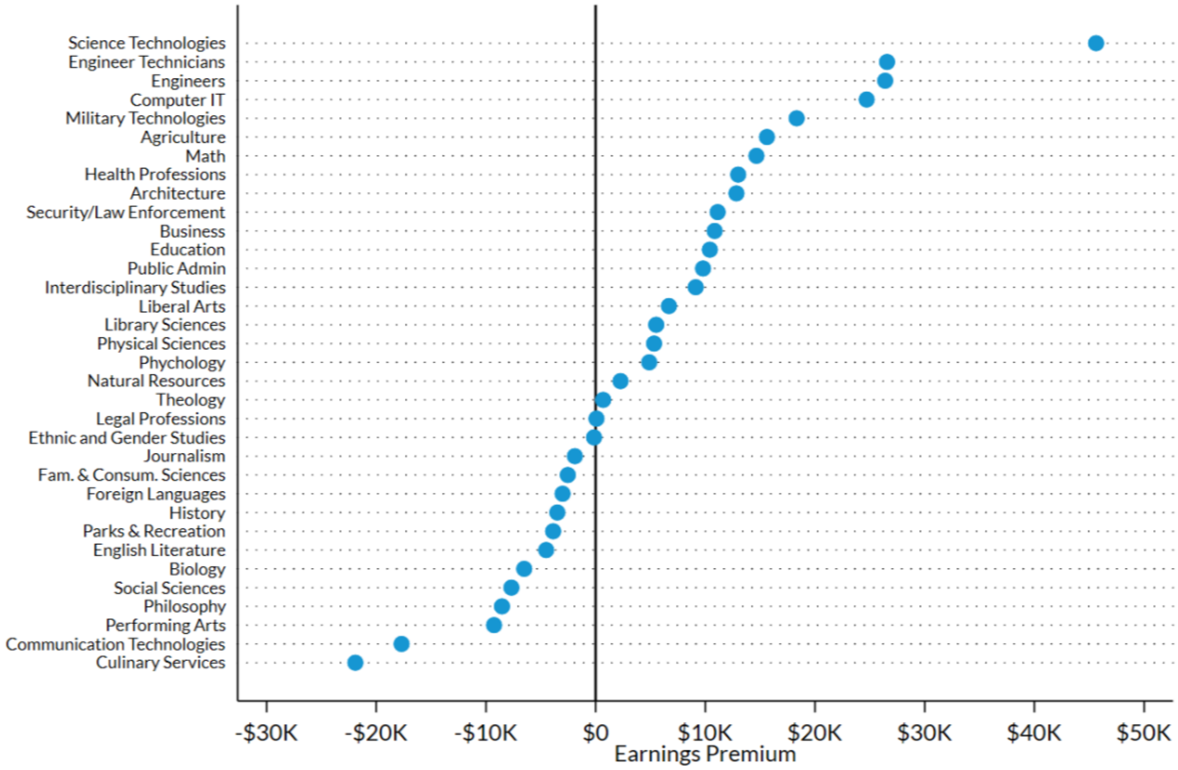


URBAN INSTITUTE

Source: Author's calculations using data from the US Department of Education and the US Census Bureau.
Notes: NET = net earnings test. This figure reports the NET pass rate for the 20 largest master's degree programs nationwide, sorted by NET pass rate. Programs are defined at the four-digit Classification of Instructional Program code level. The 20 largest programs were determined by measuring the total number of students that graduated from the program between 2014 and 2019. *N* sizes refer to all completers between 2014 and 2019 in programs with available earnings data. See the text and appendix A for details on the NET and earnings premium measures.

FIGURE B.3

Average Earnings Premiums, by Program

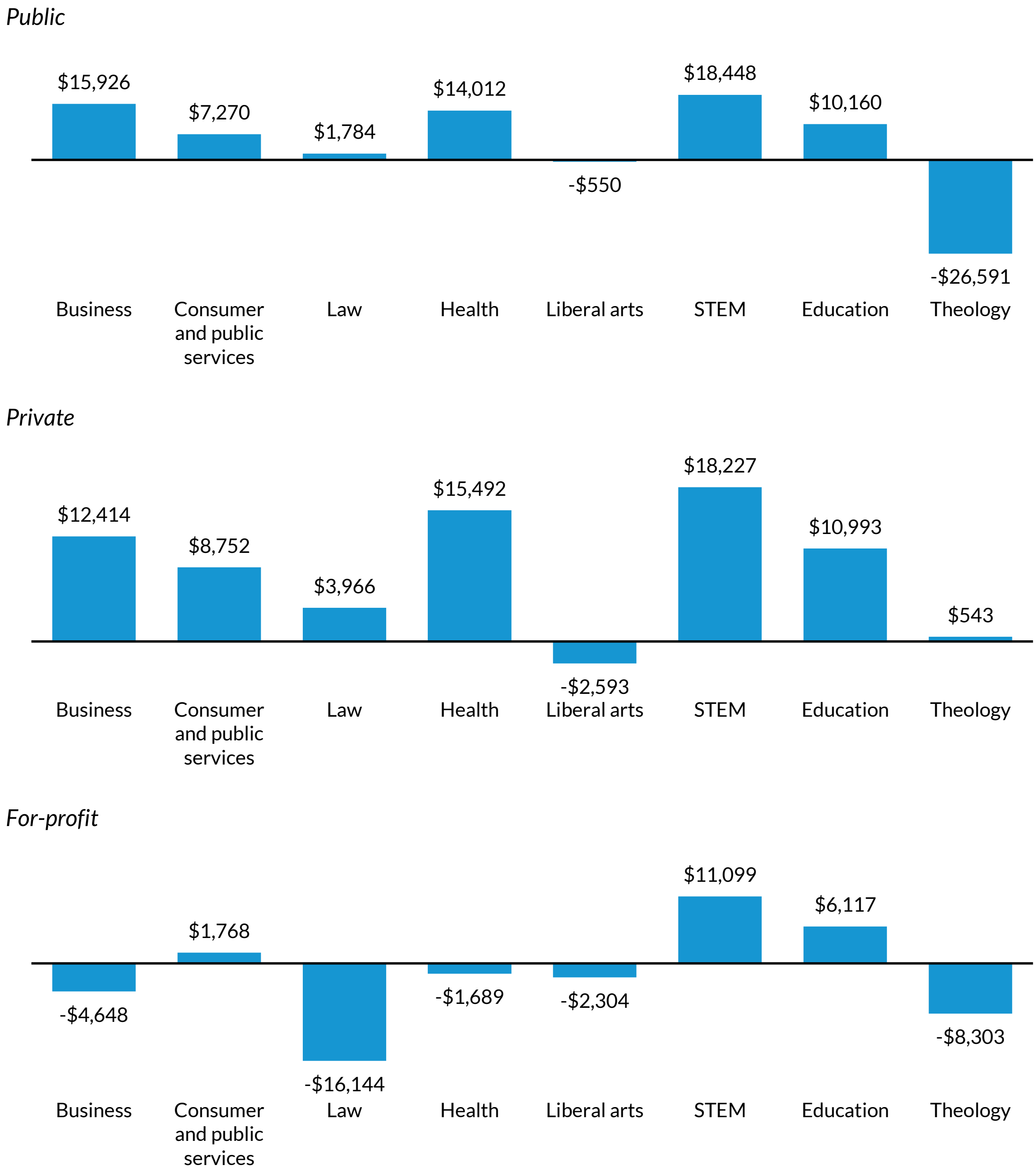


URBAN INSTITUTE

Source: Author's calculations using data from the US Department of Education and the US Census Bureau**.**
Notes: Programs are defined at the two-digit Classification of Instructional Program code level. All dollar values are in constant 2016 dollars, adjusted using the Consumer Price Index. See the text and appendix A for a description of the earnings premium.

FIGURE B.4

Earnings Premiums, by Broad Field of Study and Institutional Control



URBAN INSTITUTE

Source: Author's calculations using data from the US Department of Education and the US Census Bureau.
Notes: CIP = Classification of Instructional Programs; STEM = science, technology, engineering, and mathematics. This figure reports the unweighted average earnings premium for broad CIP groups across institution control. Programs are placed into CIP groups using the crosswalk in appendix C. All dollar values are in constant 2016 dollars, adjusted using the Consumer Price Index. See the text and appendix A for a description of the earnings premium.

Appendix C. Crosswalk for Linking CIP Codes, ACS Majors, and Field of Study Categories

TABLE C.1

CIP-ACS Broad Field of Study Crosswalk

Broad field of study category	ACS degree fields	CIP codes
Arts and humanities	Architecture; area, ethnic, and civilization studies; linguistics and foreign languages; English language, literature, and composition; liberal arts and humanities; interdisciplinary and multidisciplinary; philosophy and religious studies; theology and religious vocations; fine arts; history	Architecture and related services; area, ethnic, cultural, gender, and group studies; foreign languages, literatures, and linguistics; English language and literature and letters; liberal arts and sciences, general studies, and humanities; multi- and interdisciplinary studies; philosophy and religious studies; theology and religious vocations; visual and performing arts; history
Education and public service	Education administration and teaching; library science; psychology; criminal justice and fire protection; public affairs, policy, and social work	Education; library science; psychology; homeland security, law enforcement, firefighting, and related protective services; public administration and social service professions
Agriculture, consumer services, and trades	Agriculture; environmental and natural resources; communications; communication technologies; cosmetology services and culinary arts; family and consumer sciences; physical fitness, parks, and recreation; construction services; electrical and mechanical repairs and technologies; transportation sciences and technologies	Agricultural, animal, plant, veterinary science, and related fields; natural resources and conservation; communication, journalism, and related programs; communications technologies and technicians and support services; culinary, entertainment, and personal services; family and consumer sciences and human sciences; parks, recreation, leisure, fitness, and kinesiology; construction trades; mechanic and repair technologies and technicians; precision production; transportation and materials moving
Business and social science	Law, social sciences, business	Legal professions and studies; social sciences; business, management, marketing, and related support services
STEM and health	Computer and information sciences; engineering; engineering technologies; biology and life sciences; mathematics and statistics; military technologies; physical sciences; nuclear, industrial radiology, and biology; medical and health sciences and services	Computer and information sciences and support services; engineering; engineering and engineering-related technologies and technicians; mathematics and statistics; military science, leadership, and operational art; military technologies and applied sciences; physical sciences; science technologies and technicians; health professions and related programs; health professions residency and fellowship programs

Source: Adopted and modified from the crosswalk originally proposed in Jordan D. Matsudaira and Lesley J. Turner, *Towards a Framework for Accountability for Federal Financial Assistance Programs in Postsecondary Education* (Washington, DC: Brookings Institution, 2020).

Note: ACS = American Community Survey; CIP = Classification of Instructional Programs; STEM = science, technology, engineering, and mathematics.

Appendix D. Regression Estimates

This appendix estimates four linear regression models to examine whether the negative relationship between student demographics and earnings persists after controlling for observable program, institution, and state-level characteristics. The purpose of this analysis is to isolate how changing the composition of students within a program is associated with changes in the program's earnings outcomes while controlling for other factors that may simultaneously influence program graduates' earnings. For example, if the correlation between student characteristics and program earnings is attributable to students with particular characteristics sorting into fields with lower labor market returns, that correlation should disappear (or shrink) when controlling for a program's field of study.

Model Specification

I estimate four regression models. Each regression model iteratively adds different controls. In the baseline model (model 1), I estimate the following regression using ordinary least squares:

$$Y_i = \beta_0 + \beta_1 Pct\_Black_i + \beta_2 Pct\_Hisp_i + \beta_3 Pct\_Asian_i + \beta_4 Pct\_OtherRace_i + \beta_5 Pct\_Female_i + \epsilon_i \quad (1)$$

where Y_i is the earnings premium for master's degree program i, Pct_Black is the percentage of completers in program i that are Black, Pct_Hisp is the percentage of completers in program i that are Hispanic, Pct_Asian is the percentage of completers in program i that are Asian, $Pct\_OtherRace$ is the percentage of completers in program i that are from another race category (including race unknown, international resident, or two or more races), and Pct_Female is the percentage of completers in program i that are female. The percentage of completers in a program that are white and male are the two reference groups, respectively.

In model 2, I add covariates to control for the following institutional characteristics: a binary indicator for whether the institution is a state flagship; a binary indicator for whether the institution is classified as a Research 1 or Research 2 institution; binary indicators for institution control (public is the omitted reference); logged total 12-month full-time equivalent enrollment; logged graduate 12-month full-time equivalent enrollment; per student dollars received in federal, state, and local appropriations; per student spending on instructional salaries, instruction, research, academic supports, student services, and academic support services; and institutional endowment size at the start of the year and the end of the year.

Model 3 adds program fixed effects. Programs are defined at the four-digit CIP code level. Lastly, model 4 adds state fixed effects. In all models, robust standard errors are clustered at the institution level. For all regression models, I can control only for observable program characteristics. I cannot account for unobserved program-level characteristics. This means estimates could be biased if unobserved characteristics correlate with the outcome. Thus, I encourage caution when interpreting these results, and these descriptive results should not be interpreted as causal.

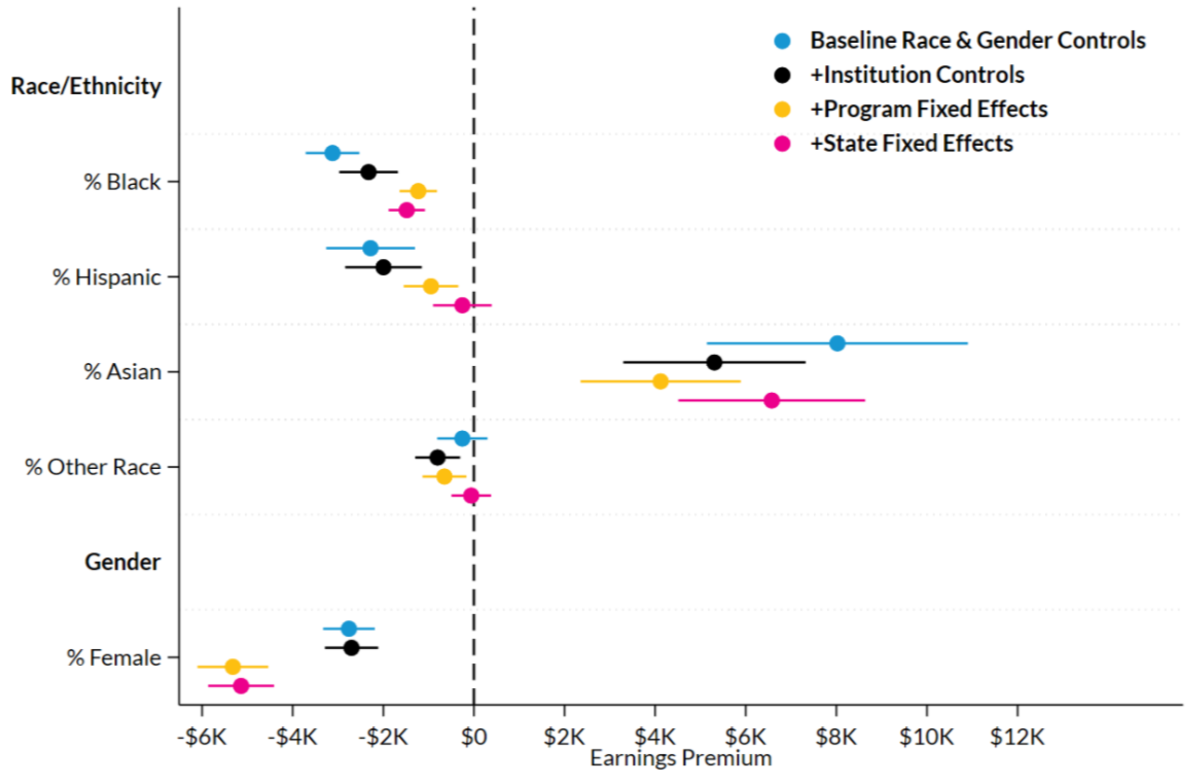
Results

Results are presented in figure D.1. In the baseline model, which accounts for only the program's race and gender characteristics, a 10 percentage-point increase in the share of Black students in a master's degree program is associated with a decline in a program's EP by $3,100 ($p < 0.001$). In other words, if Black students make up 11 percent of a master's degree program, and that share increased to 21 percent (and the white student share declined by 10 percentage points), the program's EP would be expected to drop by $3,100, on average. The size of the correlation declines when adding institutional controls, program controls, and state controls, but the relationship remains statistically significant. In the final model, which includes all institutional, program, and state controls, a 10 percentage-point increase in the Black share of students (and a 10 percentage-point decline in the share of white students) is associated with a decline in program earnings by $1,500, on average ($p < 0.001$).

The story is different for the gender composition of students within master's degree programs. In the first model, a 10 percentage-point increase in the share of students that are female (in reference to an increase in male students) is associated with a $2,800 decline in a program's EP, on average ($p < 0.001$). When also accounting for institutional characteristics, the association slightly falls to $2,700. But after accounting for a program's field of study, the association *increases* to $5,300 ($p < 0.001$). Said differently, when comparing two programs within the same field of study (e.g., two master's degree programs in public administration), all else equal, a program with a 50:50 female-male split is anticipated to have an average earnings premium $5,300 higher than a program with a 60:40 female-male split. This gap implies that gender within a program has a significant influence on the program's earnings outcomes.

FIGURE D.1

Estimated Change in Earnings Premiums Associated with a 10-Percentage Point Increase in Students with Various Race or Ethnicity and Gender Characteristics



URBAN INSTITUTE

Source: Author's calculations using data from the US Department of Education and the US Census Bureau.
Notes: This figure displays the race or ethnicity and gender regression coefficients from the four models described above. Coefficients are interpreted as the association from a 10 percentage-point increase in the X variable (i.e., the given student demographic characteristic). "% White" is the reference category for race and ethnicity variables, and "% Male" is the reference category for gender variables. Programs are weighted by the total number of completers between 2014 and 2019. Robust standard errors are clustered at the institution level. For models 3 and 4, which includes program fixed effects, programs are defined at the four-digit Classification of Instructional Program code level. Ninety-five percent confidence intervals are depicted by the horizontal lines. All dollar values are in constant 2016 dollars, adjusted using the Consumer Price Index.

Errata

This essay was corrected January 11, 2024, to include the Arnold Foundation in the acknowledgments section.

Cody L. Christensen is a doctoral student in higher education policy and leadership at Vanderbilt University.

Acknowledgments

This essay was supported by the Walton Family Foundation, the Bill & Melinda Gates Foundation, and Arnold Ventures as part of the Learning Curve essay series. 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 www.urban.org/fundingprinciples.

I am grateful for the helpful comments and feedback from Jason Delisle and Matthew Chingos. Any remaining errors are solely my own.

500 L'Enfant Plaza SW
Washington, DC 20024

www.urban.org

ABOUT THE URBAN INSTITUTE

The Urban Institute is a nonprofit research organization that provides data and evidence to help advance upward mobility and equity. We are a trusted source for changemakers who seek to strengthen decisionmaking, create inclusive economic growth, and improve the well-being of families and communities. For more than 50 years, Urban has delivered facts that inspire solutions—and this remains our charge today.